TESIS DE LA UNIVERSIDAD DE ZARAGOZA

2012

26

Jesús Ponce de León Vázquez

Análisis y síntesis de señales de audio a través de la Transformada Wavelet continua y compleja: el algoritmo CWAS

Departamento
Ingeniería Electrónica y Comunicaciones

Director/es
Beltrán Blázquez, José Ramón

ISSN 2254-7606

Prensas de la Universidad Universidad Zaragoza

Universidad Zaragoza
1542

Tesis Doctoral

ANÁLISIS Y SÍNTESIS DE SEÑALES DE AUDIO A TRAVÉS DE LA TRANSFORMADA WAVELET CONTINUA Y COMPLEJA: EL ALGORITMO CWAS

Autor

Jesús Ponce de León Vázquez

Director/es

Beltrán Blázquez, José Ramón

UNIVERSIDAD DE ZARAGOZA

Ingeniería Electrónica y Comunicaciones

2012

Departamento de
Ingeniería Electrónica
y Comunicaciones
Universidad Zaragoza

Universidad de Zaragoza
Departamento de Ingeniería Electrónica y Comunicaciones

Tesis Doctoral

Análisis y Síntesis de Señales de Audio a Través de la Transformada Wavelet Continua y Compleja: El Algoritmo C.W.A.S.

Autor: Jesús Ponce de León Vázquez

Director: José Ramón Beltrán Blázquez

Zaragoza España

Prefacio

Esta Tesis es presentada como parte de los requisitos para optar al grado académico de Doctor por la Universidad de Zaragoza, y no ha sido presentada previamente para la obtención de otro título en esta Universidad u otras. La misma contiene los resultados obtenidos en investigaciones llevadas a cabo en el Departamento de Ingeniería Electrónica y Comunicaciones, durante el período comprendido entre el 1 de noviembre de 2004 y el 31 de enero de 2012, bajo la dirección del Dr. José Ramón Beltrán Blázquez, Profesor Titular del Departamento de Ingeniería Electrónica y Comunicaciones de la Universidad de Zaragoza.

Jesús Ponce de León Vázquez
jponce@unizar.es
Departamento de Ingeniería Electrónica y Comunicaciones
Universidad de Zaragoza
Zaragoza, 7 de Mayo de 2012.

A mis padres, a los que les debo todo.
A mi familia, sin cuyo apoyo no lo habría conseguido.
A José Ramón, que nunca perderá el norte.
A Eva, por estar siempre a mi lado.

"No puedes predecir el futuro, pero puedes inventarlo".
Dennis Gabor (1900-1979), Premio Nobel de Física. *"Inventing the future"*, 1963.

Agradecimientos

Hasta hace poco, la evidente tendencia a que los *sufrientes* escribamos la palabra Tesis con "T" me parecía una simple pretensión de lucimiento, pero ahora comprendo que completar una Tesis es un camino demasiado largo y complicado para no dejar que el orgullo se trasluzca en esa "T". Llegar a poner el último punto final (que se ríe de uno mientras se le escurre entre los dedos y retrocede burlón, paso a paso) es un motivo de satisfacción personal enorme. Ser Doctor en Físicas es una de las grandes metas que había trazado en mi vida desde que vi por primera vez la serie *"Cosmos"* de Carl Sagan a los diez años más o menos, de modo que haber cumplido finalmente con ese objetivo me hace sentir tentaciones de referirme a mi Tesis con todas las letras en mayúscula. No lo haré, pero eso no significa que esté menos satisfecho con el trabajo que he realizado.

La lista de personas y entidades a las que se debe en parte o en su totalidad la elaboración de esta Tesis, es interminable. De modo que si alguien se echa de menos, acepte mi disculpa por anticipado.

En el terreno puramente profesional, quisiera mostrar mi agradecimiento a la Universidad de Zaragoza, al Ministerio de Educación y Ciencia y a los demás organismos que han financiado en su totalidad o en parte el trabajo de investigación que se ha llevado a cabo.

A nivel personal, me gustaría en primer lugar agradecer a mis padres el esfuerzo que han invertido en conseguir que mi formación llegase hasta donde lo ha hecho. *Esta tesis es el resultado final de vuestro apoyo.* Mis hermanos, sobrinos, cuñados, suegros y demás familia han acabado por acostumbrarse a tener un científico en casa todos estos años. *Gracias a todos, y recordad que "¡aquí falla algo!".* Durante casi toda la elaboración de este trabajo, mi mujer ha soportado estoicamente las complicadas situaciones profesionales que he atravesado sin proferir una queja, compartiendo conmigo un único deseo: la finalización de esta Tesis. Bueno, pues ya está hecho. Quien sabe, ¡igual a partir de ahora conseguimos algunas vacaciones sin el portátil en la mochila! *Decir gracias es poco, guapetona. Nunca lo hubiese conseguido sin tu ayuda.*

No puedo dejar de agradecer lo que me dieron cada uno de aquellos pedagogos por cuyas manos he pasado en algún momento de mi vida. De todos guardo gratísimos recuerdos, aunque me limitaré a nombrar a los más entrañables: Don Paco, Don Miguel, Doña Francisca, Don Armando, Mr. Fernando, Chema, Carmen, César, Pilar, José Carrasquer... Cada uno de ellos puso una pieza de lo que soy. *Gracias a todos, estéis donde estéis.*

He pasado varios años deambulando por los pasillos de la Facultad de Ciencias, la Escuela de Ingenieros y el Centro Politécnico Superior. En todos y cada uno de estos lugares, como alumno primero y como profesor después, he conocido gente por la que siento la más

profunda admiración. He aquí unos pocos de los destacados: mis profesores (algunos de ellos ya desaparecidos) Manuel Quintanilla, José María Sabirón, Luis Joaquín Boya, Manuel Asorey, Ángel Morales, Pepe Barquillas, y Pedro Martínez, entre otros, que me han enseñado a sufrir, comprender y amar el siempre fascinante mundo de la física, las matemáticas y la electrónica. De entre mis compañeros de profesión, la lista de personas por las que mi afecto y agradecimiento destaca, resulta especialmente selecta. Mis camaradas de la antigua EUITIZ, Vicente, Boni, Antonio, Carlos; mis colegas del desaparecido CPS, Estanis, Fernando, Armando, Abelardo. *Sois todos colosales.*

Una mañana se abrió la puerta de mi despacho y asomó la cabeza por el umbral un hombre con el que hasta entonces apenas me sentía unido por un tenue lazo de sintonía (ambos somos físicos) y por una asignatura de Teleco heredada (Instrumentación Electrónica). Yo era el *nuevo*, recién llegado de tres poco fructíferos años de profesor de prácticas en la Facultad de Ciencias. Ese hombre podía haberme ignorado sin más, y sin embargo aquella mañana pronunció una frase que llevaba casi cuatro años esperando oír: *"¿Quieres hacer una tesis doctoral conmigo?"*. Mi respuesta fue inmediata: *"Sí"*. Se trata de José Ramón Beltrán, mi director y mi compañero de trabajo, pero sobre todo mi amigo. No tengo palabras para agradecer tu esfuerzo, paciencia y dedicación. En todos los aspectos has demostrado ser comprensivo a la par que eficiente, sin dejar por ello de exigir lo mejor de mí. Has sabido darme aliento en los períodos en que he flaqueado, y ambos sabemos que han sido numerosos. Esta Tesis es tan tuya como mía. Espero tener la oportunidad de reponer la energía que has derrochado conmigo. *Gracias por todo.*

Por último, quisiera recordar a un compañero que se fue, el cual me demostró que, al contrario de lo que tan a menudo parece alentarse a nuestro alrededor, amar la docencia no es pusilánime: mi amigo Tomás Pollán. *Espero parecerme un poco a ti algún día, Tomás.*

Jesús Ponce de León Vázquez
jponce@unizar.es
Departamento de Ingeniería Electrónica y Comunicaciones
Universidad de Zaragoza
Zaragoza, 7 de Mayo de 2012.

Resumen

En esta Tesis se pretende demostrar que la Transformada Wavelet Continua y Compleja (CCWT) puede ser una herramienta precisa para la obtención de características de alto nivel de la señal de audio, a través de un algoritmo generalista del modelo que se propone de la misma.

En los siguientes seis capítulos se va a presentar un algoritmo funcional basado en la CCWT, el algoritmo de Síntesis Aditiva por Wavelet Complejas, o CWAS por sus siglas en inglés (Complex Wavelet Additive Synthesis). El algoritmo CWAS parte de una versión previa presentada por José Ramón y Fernando Beltrán en 2003, [17]. Como se verá en los Capítulos 1 y 2, este algoritmo inicial presentaba una serie de problemas que parecían no encajar con el marco teórico que se desprende de la literatura.

Para revelar el origen de estos problemas, se hace necesario un análisis matemático riguroso de los coeficientes wavelet (Capítulo 2). A partir de algunas ideas introducidas en el desarrollo matemático presentado, la solución para las limitaciones de la técnica inicial ha permitido llegar finalmente la obtención de un novedoso modelo de la señal de audio (Capítulo 3) bien posicionado de cara a posibles aplicaciones. Como se verá en los Capítulos 2 y 3, un filtrado pasobanda complejo unitario permite el cálculo de los coeficientes wavelet, en cuyo módulo se indican de forma implícita las bandas del espectro de frecuencia que conforman la *zona de influencia* de cada componente detectada. La suma de los coeficientes wavelet en las bandas asociadas a cada componente proporciona una función compleja para cada parcial, de amplitud y fase instantáneas altamente coherentes (es decir, muy cercanas al par canónico teórico de la señal). Es precisamente la coherencia en fase la principal ventaja del modelo propuesto. Un simple modelo se síntesis aditiva permite la generación de una señal sintética de características tímbricas y tonales muy similares a la señal original, con la característica añadida de una diferencia punto por punto entre las señales analizada y sintética que resulta despreciable numéricamente para la mayoría de las aplicaciones.

El modelo subyacente permite colocar al algoritmo CWAS en buena posición de cara a distintas aplicaciones. De hecho, se ha utilizado en síntesis de sonidos, localización de onsets y detección de fundamentales entre otras aplicaciones, con resultados muy prometedores (Capítulos 4 y 5). Del mismo modo, se han hecho grandes progresos de cara a aplicaciones

más completas y complejas, como la separación ciega de fuentes monaurales de sonido. En este caso, se han desarrollado tres técnicas de complejidad creciente, la más completa de las cuales será detallada en el Capítulo 4. Esta técnica se basa en la reconstrucción completa de la información de los parciales superpuestos (amplitud y fase) a partir de los parciales aislados, y ofrece unos resultados de gran calidad sonora y numérica. Sin embargo, aún no se ha llegado a desarrollar un algoritmo completo de separación de fuentes. Como se verá en el Capítulo 4, en su estado actual, el algoritmo es un buen separador de notas musicales. Otras aplicaciones desarrolladas pueden verse en el Apéndice II, entre las cuales cabe destacar dos técnicas de separación adicionales.

El algoritmo CWAS presenta una serie de ventajas e inconvenientes sobre otras técnicas basadas en diferentes Distribuciones Tiempo–Frecuencia, como la STFT (Capítulo 5). Entre las ventajas, a partir de la coherencia en fase se consiguen resultados en la síntesis de sonido por encima de los arrojados utilizando otras técnicas, lo cual permite abordar con elevadas esperanzas de éxito aplicaciones más ambiciosas que las aquí presentadas, como un algoritmo más completo y eficiente de separación de señales monaurales. La principal limitación del algoritmo propuesto es el tiempo de procesado, que ha impedido por el momento el empleo de esta técnica en aplicaciones en tiempo real. Sin embargo, como se demostrará, el algoritmo CWAS resulta ser sensiblemente más rápido que la STFT en condiciones de trabajo equivalentes, y la linea de investigación hacia este tipo de aplicaciones está abierta y activa.

El algoritmo CWAS y todas sus aplicaciones, así como todas las gráficas (a excepción de los diagramas algorítmicos, generados con Microsoft® Office Powerpoint® 2007) y datos numéricos presentados en este trabajo, han sido desarrollados y generados por el autor en entorno Matlab®, versiones 6.5 a 7.8.0.347, salvo indicación expresa de lo contrario. A lo largo del texto, se remitirá con cierta frecuencia a la calidad de las señales sintetizadas, filtradas o separadas. El presente documento viene acompañado por un soporte digital en el que se almacenan la mayoría de los resultados sonoros aludidos en los siguientes capítulos y anexos.

Abstract

This Thesis aims to show that the Complex Continuous Wavelet Transform (CCWT) can be an accurate tool for obtaining high-level features of the audio signal, through a general algorithm of the proposed model.

In the following six chapters, a functional algorithm based on the CCWT will be presented: the Complex Wavelet Additive Synthesis or CWAS algorithm. This algorithm comes from a previous version performed by José Ramón and Fernando Beltrán in 2003, [17]. As discussed in Chapters 1 and 2, the initial algorithm presented certain problems that seemed not to fit with the theoretical framework that emerges from the literature.

In order to reveal the origin of these problems, a rigorous mathematical analysis of the wavelet coefficients is necessary (Chapter 2). Based on some of the ideas introduced in the presented mathematical development, the solution to the limitations of the initial technique finally allowed the obtention of a new model of the audio signal (Chapter 3), well positioned to face possible applications. As discussed in Chapters 2 and 3, a unitary complex bandpass filtering allows the calculation of the wavelet coefficients, in which module they are implicitly indicated the bands of frequency of the spectral shape which define the *zone of influence* of each detected component. The summation of the wavelet coefficients in the bands associated with each component provides a complex function for each partial, which presents highly coherent instantaneous amplitude and phase (that is, close to the theoretical canonical pair of the signal). The most important advantage of the proposed model is precisely the phase coherence. A simple additive synthesis model allows to generate a synthetic signal which timbre and pitch characteristics very similar to the original signal. It also allows to calculate a *point-by-point* difference between analyzed and synthetic signals, which results to be negligible for most applications.

The underlying model allows to place the CWAS algorithm in good position to confront different applications. In fact, it has been used in sound synthesis, localization of onsets and fundamental frequency estimation among other applications, always with promising results (Chapters 4 y 5). Similarly, some great progress has been made towards more complex applications, such as Blind Audio Source Separation of monaural signals (Chapter 4). In this case, we have developed three techniques of increasing complexity, the most complete

of which is detailed in Chapter 4. This technique is based on the complete reconstruction of the information (amplitude and phase) of the overlapping partials using the envelopes and phases of the isolated partials. It offers high quality acoustical and numerical results. However, we have not yet developed a full algorithm of source separation. As discussed in Chapter 4, in its current state, the algorithm is a good separator of musical notes. Other developed applications are shown in Appendix II, among which we include two additional separation techniques.

The CWAS algorithm presents some advantages and disadvantages with respect to other techniques, based on different Time-Frequency Distributions (as the STFT, Chapter 5). Among the advantages, results achieved in the synthesis of sound (related with the phase coherence) are above the obtained results using other techniques. This allows confronting, with high hopes of success, to more ambitious applications than the applications presented in this dissertation, as for example a complete and efficient algorithm of blind separation of monaural signals. The main limitation of the proposed technique is the processing time, which currently prevents the use of this technique in real-time applications. However, as will be shown, the CWAS algorithm is quite faster than the STFT under equivalent working conditions.

The CWAS algorithm and all its applications, as well as al the displayed figures and graphics (except the algorithmic blocks, generated using Microsoft® Office Powerpoint® 2007) and numerical data presented in this work, have been developed by the author in a Matlab® context (versions 6.5 to 7.8.0.347), unless specified. Throughout the text, the quality of the synthesized signals will be cited with some frequency. The present document is accompanied by a CD that stores most of the sounds cited in in the following chapters and appendixes.

Índice general

Índice de figuras

Índice de tablas

Capítulo 1

Estado del Arte y Objetivos Generales

Índice

"¡Qué maravilloso habernos
encontrado con una paradoja!
Ahora tenemos alguna esperanza
de avanzar".
Niels Henrik David Bohr (1885-1962).
Físico danés.

En el presente capítulo, se revisa con cierto nivel de detalle la obtención de la evolución temporal de la información energética y frecuencial de la señal de audio. Asimismo, se explicarán algunas de las más destacadas técnicas de análisis existentes y sus limitaciones, y en particular de la Transformada Wavelet, que es la herramienta utilizada y desarrollada en el presente trabajo, y que puede ser utilizada en alguna de sus variantes en múltiples disciplinas, desde el análisis de señales biomédicas al filtrado y procesado de imágenes. Para finalizar, se expondrán los objetivos generales de la presente disertación, así como su estructura final.

1.1. Introducción

En esta Tesis se va a introducir, desarrollar y testear un algoritmo que toma como base la Transformada Wavelet Continua y Compleja (CCWT), como método de análisis/síntesis de señales no estacionarias. El algoritmo, que será detallado principalmente a lo largo del Capítulo 3 y sus aplicaciones (Capítulos 4 y 5), han sido desarrollados completamente en entorno Matlab®.

Para comenzar, en el presente capítulo se introducen una serie de conceptos básicos que serán utilizados en adelante, como el de *frecuencia instantánea* o el de *señal analítica*, así como otros conceptos no menos interesantes, como el de *par canónico*. A continuación se introducen algunas de las transformaciones tiempo–frecuencia más empleadas en el análisis de señales no estacionarias, entre las que cabe destacar tres: la Transformada de Fourier Localizada, la Distribución de Wigner-Ville y por último la Transformada Wavelet.

Una vez introducida esta última transformada, se procederá a detallar una serie de definiciones matemáticas y ecuaciones características básicas de la misma, así como las particularidades de la Transformada Wavelet Continua y Compleja, y se introducirán conceptos como el de *cresta* y *esqueleto*, llevándose a cabo un repaso general de la literatura que se ha desarrollado en las tres últimas décadas acerca de esta importante rama del análisis de señal.

Los resultados presentados en esta Tesis están dirigidos principalmente al análisis y síntesis de señales de audio, si bien la CCWT es potencialmente utilizable en otras muchas aplicaciones. Algunas de estas aplicaciones (las cuales emplean la Transformada Wavelet real o compleja, en una o varias dimensiones) serán introducidas en la parte final de este estudio del estado del arte.

Para finalizar, se definen los objetivos generales y puntuales de la presente disertación, así como la estructura general de la misma.

1.2. Conceptos: Frecuencia Instantánea y Par Canónico

Antes de comenzar a desarrollar el tema de las Representaciones Tiempo–Frecuencia, conviene asentar adecuadamente algunos conceptos de uso común que resultarán muy importantes durante el resto de este trabajo, como el de *frecuencia instantánea*, *señal analítica* y *par canónico*, entre otros.

La complejidad de la frecuencia instantánea, $f_{ins}(t)$, comienza por su propio nombre: el calificativo *instantáneo* (efímero, breve; o, como lo define la Real Academia Española de la Lengua[1]: *"Que solo dura un instante"*) parece no encajar con el concepto clásico de *frecuencia* (que implica evolución continua, periodicidad). La frecuencia instantánea de una señal monocromática, puede ser entendida como un parámetro univaluado, variable en el tiempo, que sigue la localización espectral de su única componente, a medida que ésta evoluciona. En ocasiones se puede interpretar como la frecuencia promedio de la señal en cada instante de tiempo [124]. En [31] se presenta un trabajo bastante completo acerca de las diferentes interpretaciones físicas y definiciones matemáticas que rodean al amplio concepto de frecuencia instantánea.

Probablemente, la definición de frecuencia instantánea más cargada de sentido físico proviene de la definición de *fase instantánea*. En su expresión más general, una señal puede escribirse como:

$$x(t) = A(t)\cos[\phi(t)] \tag{1.1}$$

En esta ecuación, $A(t)$ es la amplitud o envolvente instantánea de la señal, mientras que $\phi(t)$ es su fase instantánea. En el caso de una onda de amplitud constante y frecuencia pura, en la cual $A(t) = A_1$ y $\phi(t) = \omega_1 t$, tanto la amplitud como la frecuencia instantáneas son conceptos claros e intuitivos. Cuando $x(t)$ es no estacionaria, tanto $A(t)$ como $\phi(t)$ son funciones dependientes del tiempo. En tal situación, y en el caso más general, la frontera entre amplitud y frecuencia instantánea no está muy clara. Es posible distinguir entre ambos conceptos siempre que se pueda separar suficientemente sus espectros asociados. Esta condición se conoce en la literatura como *aproximación asintótica*, y puede expresarse

[1] Real Academia Española. *Diccionario de la Lengua Española*, Vigésima edición, 1984.

matemáticamente [48, 53] como:

$$\left|\frac{1}{A(t)}\frac{d\ A(t)}{dt}\right| \ll \left|\frac{d\phi(t)}{dt}\right| \tag{1.2}$$

Es decir, cuanto mayor sea la distancia que separa el espectro de $A(t)$ del de $\phi(t)$, mejor describe la Ecuación (1.1) a $x(t)$ [15, 30, 131]. La parte inferior del espectro se corresponde a la amplitud instantánea u onda *moduladora*, $A(t)$, mientras que las frecuencias más elevadas son para la fase instantánea u onda *portadora*, $\phi(t)$.

La representación descrita por la Ecuación (1.1) está lejos de ser única para $x(t)$. De hecho, existen infinitos pares de funciones $\big(A(t),\ \phi(t)\big)$ que satisfacen la misma ecuación, y por lo tanto se relacionan con la misma señal. Esta ambigüedad puede evitarse obteniendo lo que se llama el *par canónico* de la señal, $\big(A_x \geq 0,\ \phi_x \epsilon [0, 2\pi]\big)$ [131], el cual a su vez está relacionado con la *señal analítica* relativa a $x(t)$ [67], $x_{an}(t)$. En efecto, cada función expresada por la Ecuación (1.1) tiene asociada una *única* señal analítica. A este resultado se conoce como el *Teorema del Producto de Bedrosian* [15]. La señal analítica $x_{an}(t)$ se calcula partiendo de $x(t)$ a través de la Transformada de Hilbert (HT) [31, 15, 77]. Concretamente:

$$x_{an}(t) = A_x(t)\cos[\phi_x(t)] + jHT\{A_x(t)\cos[\phi_x(t)]\} \tag{1.3}$$

El problema de la resolución general de esta ecuación ha sido ampliamente estudiado (véase, por ejemplo, [136]). Aplicando el Teorema del Producto de Bedrosian se deduce que, siempre que el espectro en frecuencia de la moduladora, $A(t)$, esté completamente incluido en la región de frecuencia $f < f_0$ y el espectro de la portadora exista exclusivamente fuera de esta región (es decir, en tanto en cuanto la Ecuación (1.2) se cumpla), la Transformada de Hilbert se limita aproximadamente a introducir un desfase de 90° en el coseno de la Ecuación (1.3), resultando:

$$HT\{A_x(t)\cos[\phi_x(t)]\} \approx A_x(t)\sin[\phi(t)] \tag{1.4}$$

y por lo tanto:

$$x_{an}(t) \approx A_x(t)\cos[\phi_x(t)] + jA_x(t)\sin[\phi(t)] = A_x(t)e^{j\phi_x(t)} \tag{1.5}$$

donde e es la base del logaritmo natural y j es la variable compleja.

La aproximación asintótica de la Ecuación (1.2) está a su vez relacionada con la *duración efectiva* de la señal, T, y su *ancho de banda efectivo*, B [31]. La duración efectiva (T) se define como el período de tiempo durante el cual la amplitud (energía) de $x(t)$ está por encima de cierto umbral, como por ejemplo el impuesto por el ruido ambiental. El ancho de banda efectivo (B) se define de forma equivalente, si bien en el dominio frecuencial. Las

señales asintóticas se caracterizan por tener un producto BT suficientemente grande.

Para cada $x(t)$, sus relativos T_x y B_x pueden ser expresados [67, 31, 156] como:

$$T_x^2 = \frac{\int_{-\infty}^{+\infty} t^2|x(t)|^2dt}{\int_{-\infty}^{+\infty} |x(t)|^2dt} \tag{1.6}$$

$$B_x^2 = \frac{\int_{-\infty}^{+\infty} f^2|\hat{x}(\omega)|^2d\omega}{\int_{-\infty}^{+\infty} |\hat{x}(\omega)|^2d\omega} \tag{1.7}$$

donde, por supuesto, $\hat{x}(\omega)$ es la transformada de Fourier de $x(t)$.

Como se demuestra en [157], el 99 % de la energía de la señal está concentrada dentro del marco limitado por B y T si:

$$BT \geq 5 \tag{1.8}$$

Cabe destacar que la condición asintótica resulta primordial para poder distinguir entre la onda moduladora y la portadora *únicamente en el sentido físico*. Pero incluso si la Ecuación (1.2) no es cierta, tanto la amplitud instantánea como la fase instantánea de la señal analítica continúan estando bien definidas.

Cuando $A(t)$ y $\phi(t)$ muestran espectros separados, la frecuencia instantánea $f_{ins}(t)$ de la señal tiene un significado físico muy aproximado al concepto clásico de frecuencia. Aunque se puede definir de diversas formas, quizá la expresión matemática más útil para $f_{ins}(t)$ [31, 51, 76] es:

$$f_{ins}(t) = \frac{1}{2\pi}\frac{d\phi(t)}{dt} \tag{1.9}$$

La desviación estándar de la frecuencia instantánea en un instante de tiempo determinado puede interpretarse como un *ancho de banda instantáneo*, IB, el cual depende básicamente de la variación frecuencial de la señal y de su producto BT. Como se indica en [31], el ancho de banda instantáneo de una señal de frecuencia instantánea $f_{ins}(t)$, analizada bajo una Distribución Tiempo–Frecuencia genérica $\rho(t, f)$ sería:

$$IB_f^2(t) = \frac{\int_{-\infty}^{\infty}[f - f_{ins}(t)]^2\rho(t, f)df}{\int_{-\infty}^{\infty}\rho(t, f)df} \tag{1.10}$$

Un concepto equivalente a este, basado exclusivamente en criterios energéticos, se empleará más adelante en este trabajo (concretamente en la Sección 3.5.1), para caracterizar los *parciales* de la señal subyacentes al modelo presentado.

El espectro en frecuencia de las señales multi–componente se localiza, para cierto instante de tiempo, alrededor de $N \geq 2$ valores diferentes de frecuencia. Por lo tanto tales señales pueden describirse como la suma de N componentes o parciales, cada uno descrito por la

Ecuación (1.1):

$$x(t) = A(t)\cos[\phi(t)] = \sum_{n=1}^{N} A_n(t)\cos[\phi_n(t)] \tag{1.11}$$

La descomposición de una onda multi–componente en sus parciales asociados no es única, depende de la aplicación [31]. Para tales señales consideradas como un conjunto, la onda moduladora $A(t)$, la portadora $\phi(t)$ e incluso la frecuencia instantánea $f_{ins}(t)$ pueden perder completamente su significado físico [125]. Sin embargo, si las Ecuaciones (1.2) ó (1.8) se cumplen *para cada componente*, cada uno de estos presenta a su vez su propia amplitud y fase instantáneas, $A_n(t)$ y $\phi_n(t)$, y por lo tanto está ligado a su *parcial analítico*. De este modo, la señal completa puede definirse como la parte real de los parciales analíticos sumados:

$$x(t) = \Re\left[\sum_{n=1}^{N} A_{n,x}(t)e^{j\phi_{n,x}(t)}\right] \tag{1.12}$$

Esta importante ecuación ofrece una relación entre el par canónico de la señal y sus correspondientes componentes. La Ecuación (1.11) es la expresión más usual del clásico modelo de *síntesis aditiva* [139, 140]. Significa, básicamente, que es posible obtener la señal analítica (y por lo tanto el par canónico) de cada componente de la señal original. De este modo también se obtiene una expresión del par canónico de la propia señal completa $x(t)$:

$$x(t) = A_{x_N}(t)\cos[\phi_{x_N}(t)] \tag{1.13}$$

donde:

$$A_{x_N}(t)^2 = A_{x_{N-1}}(t)^2 + A_{N,x}(t)^2 + 2A_{x_{N-1}}(t)A_{N,x}(t)\cos[\Delta_N(t)] \tag{1.14}$$

$$\phi_{x_N}(t) = \arctan\left[\frac{A_{x_{N-1}}(t)\sin[\Delta_N(t)]}{A_{N,x}(t) + A_{x_{N-1}}(t)\cos[\Delta_N(t)]}\right] + \phi_{N,x}(t) \tag{1.15}$$

y

$$\Delta_N(t) = \phi_{x_{N-1}}(t) - \phi_{N,x}(t) \tag{1.16}$$

En estas expresiones recursivas, $A_{N,x}$, $A_{N-1,x}$ y demás, son las amplitudes instantáneas de las diferentes componentes de $x(t)$, $\phi_{x,N}$, $\phi_{N-1,x}$,... etc., sus fases instantáneas, A_{x_N}, la envolvente global de los N componentes de la señal ($A_{x_{N-1}}$, la envolvente instantánea de las primeras $N-1$ componentes) y ϕ_{x_N} la fase global de la señal ($\phi_{x_{N-1}}$ la fase global de los primeros $N-1$ términos de la señal original).

En el caso particular de dos componentes, estas fórmulas pueden reescribirse como sigue:

$$A_{x_2}(t)^2 = A_1(t)^2 + A_2(t)^2 + 2A_1(t)A_2(t)\cos[\Delta(t)] \tag{1.17}$$

$$\phi_{x_2}(t) = \arctan\left[\frac{A_1(t)\sin[\Delta(t)]}{A_2(t) + A_1(t)\cos[\Delta(t)]}\right] + \phi_2(t) \tag{1.18}$$

y:

$$\Delta(t) = \phi_1(t) - \phi_2(t) \tag{1.19}$$

En el contexto del análisis pasobanda de la señal multi–componente, sería muy interesante encontrar el par canónico de cada componente *aislada* de $x(t)$. Para ello emplearemos el Análisis de Fourier mientras sea posible, y las técnicas de Representación Tiempo–Frecuencia allí donde no lo sea. Por otro lado, en el contexto de la percepción del sonido, no resulta estrictamente necesario separar cada componente de la señal original, si no sólo el par canónico de la misma *en las bandas espectrales que pueden ser resueltas por el oído humano*, [13, 56]. Como se verá más adelante, este hecho facilita enormemente la generación del banco de filtros que emplea el algoritmo presentado en esta Tesis, el algoritmo de *Síntesis Aditiva por Wavelets Complejas* (Complex Wavelet Additive Synthesis, C.W.A.S.) y explica muchas de sus características esenciales en cuanto a calidad final de la resíntesis.

1.3. Espectros estáticos: Análisis de Fourier

Como se ha explicado en la sección anterior, las señales estáticas son aquellas cuyo contenido frecuencial no varía en el tiempo. Para extraer la información frecuencial de una señal de este estilo, es suficiente aplicar la Transformada de Fourier. La transformada de Fourier de una función $x(t)$ se define matemáticamente como:

$$\widehat{X}(\omega) = \int_{-\infty}^{+\infty} x(t) e^{j2\pi\omega t} dt \tag{1.20}$$

Para que esta ecuación pueda aplicarse, es necesario que la función $x(t)$ sea una función (de variable real o compleja) integrable al menos una vez en el sentido de Lebesgue, es decir:

$$x \in L^1(\mathbb{R}) \qquad o \qquad x \in L^1(\mathbb{C}) \tag{1.21}$$

El resultado de aplicar la transformada de Fourier a una señal monocromática de amplitud constante y frecuencia pura, es una delta de Kronecker modulada por la amplitud de la señal y localizada en dicha frecuencia. Es decir, si la señal a analizar es , por ejemplo, el coseno de una frecuencia pura ω_1:

$$x(t) = A_1 \cos(\omega_1 t) \tag{1.22}$$

entonces su Transformada de Fourier resulta:

$$\widehat{X}(\omega) = \int_{-\infty}^{+\infty} A_1 \cos(\omega_1 t) e^{-j2\pi\omega t} dt = A_1 \delta(\omega - \omega_1) \tag{1.23}$$

De un modo equivalente, es posible conocer la evolución temporal de una señal cuyo contenido frecuencial (supuesto estático) resulte conocido, sin más que aplicar la Transformada Inversa de Fourier:

$$F^{-1}(\widehat{X}) = x(t) = \int_{-\infty}^{+\infty} \widehat{X}(\omega) e^{j2\pi\omega t} dt \tag{1.24}$$

La única diferencia entre la Transformada de Fourier y su inversa es el signo negativo en el exponente del integrando. Una de las características destacables de la Transformada de Fourier es que se trata de una función de soporte infinito. Esto quiere decir que para que una señal sinusoidal pura dé como resultado en su transformada de Fourier una delta pura es necesario que la señal a analizar tenga una duración infinita hacia delante y hacia atrás en el tiempo. Evidentemente, las señales reales no poseen esta característica, y el resultado es lo que se conoce como efectos de borde: en el caso de una señal de frecuencia constante determinada, la aparición en el espectro de la señal de componentes sinusoidales espurios, armónicos naturales de la frecuencia a detectar.

En el mundo del procesado digital de la señal, donde se trabajará con señales discretas en lugar de trabajar con la Ecuación (1.20), se emplea como herramienta de análisis la llamada Transformada Discreta de Fourier (DFT). Una secuencia de n números complejos $x_0, ..., x_{n-1}$ se transforma en la secuencia de n números complejos $f_0, ..., f_{n-1}$ mediante la DFT según la fórmula:

$$f_m = \sum_{k=0}^{n-1} x_k e^{-j\frac{2\pi}{n}km}, \qquad \forall m = 0, 1, \ldots, n-1 \tag{1.25}$$

La Transformada de Fourier Discreta Inversa (IDFT), puede obtenerse a su vez a través de la expresión:

$$x_k = \frac{1}{n} \sum_{m=0}^{n-1} f_m e^{-j\frac{2\pi}{n}mk}, \qquad \forall k = 0, 1, \ldots, n-1 \tag{1.26}$$

Nótese que los factores de normalización (1 y $1/n$, respectivamente) pueden diferir en otras definiciones de la Transformada de Fourier Discreta. Un factor de normalización de $1/n^{1/2}$ (la única condición *sine qua non* para los mismos es que su producto valga $1/n$), tanto para la Transformada Directa como para la Inversa, hace las transformaciones *unitarias* (es decir, conservan la energía de la señal analizada), lo que presenta en ocasiones ciertas ventajas.

La Transformada Discreta de Fourier puede calcularse de modo muy eficiente mediante algoritmos FFT. Tales algoritmos ponen algunas limitaciones en la señal de entrada y en el espectro resultante. Por ejemplo: la señal que se va a transformar debe consistir en un número de muestras igual a una potencia de dos (la mayoría de los algoritmos FFT, y en concreto el utilizado en la realización de este trabajo, permiten la transformación de 512,

1024, 2048 ó 4096 muestras). El rango de frecuencias cubierto por el análisis FFT depende tanto de la cantidad de muestras de la señal de entrada como de la frecuencia de muestreo.

1.4. Espectros dinámicos: representaciones Tiempo – Frecuencia

Las limitaciones más importantes de la Transformada de Fourier (ya sea continua o discreta), así como las de la mayoría de las herramientas adecuadas para analizar señales estacionarias, resultan evidentes cuando se intenta analizar una señal cuyo espectro varía con el tiempo. Por ejemplo, las señales de la parte izquierda de la Figura 1.1, arrojan resultados muy similares en sus espectros estáticos de Fourier (parte derecha de la figura), cuando es evidente que se trata de señales muy distintas. El motivo es que la transformada simple de Fourier con la que se detectan las componentes presentes en la señal bajo análisis, no ofrece por sí misma ninguna información acerca de la evolución temporal de tales componentes.

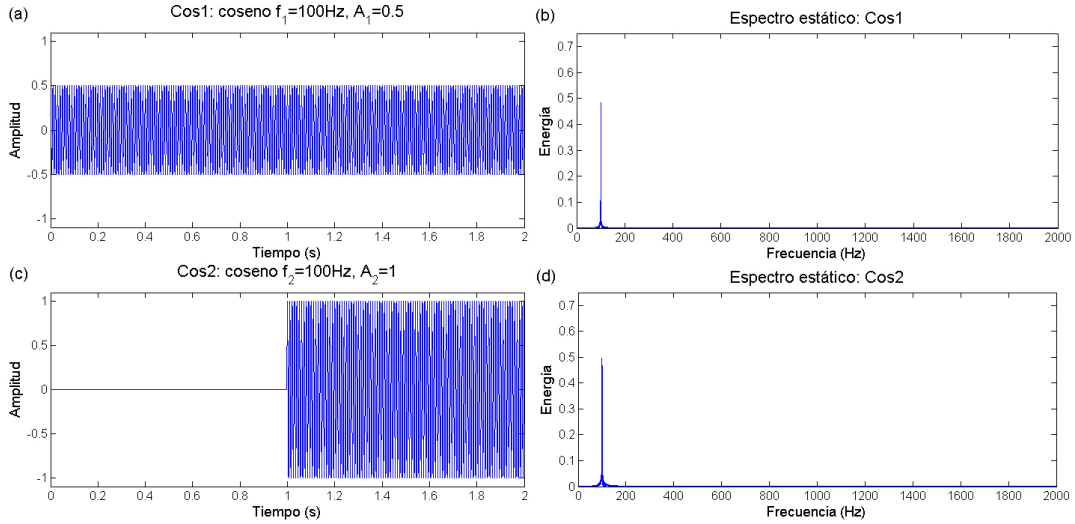


Figura 1.1: *(a) Forma de onda para coseno de 100Hz de frecuencia y 0.5 de amplitud pico a pico, con una duración total de 2 segundos. (b) Espectro estático del coseno representado en (a). (c) Forma de onda para coseno de 100Hz de frecuencia y 1 de amplitud pico a pico, con una duración total de 1 segundo. (d) Espectro estático del coseno representado en (c).*

Las señales de la Figura 1.1 están preparadas para arrojar espectros similares, puesto que su contenido energético es básicamente idéntico (doble duración, mitad de amplitud de oscilación) y su frecuencia es la misma (100Hz). Si se desea obtener una información más

exhaustiva de estas señales de espectro variable, es necesario recurrir a transformaciones que presenten de forma simultánea el contenido frecuencial de una señal y su evolución temporal, de modo que se hablará de *espectros* y *espectrogramas*, para distinguir entre el contenido frecuencial de una señal y la evolución temporal del mismo.

1.4.1. Transformaciones tiempo–frecuencia: marco histórico

El tratamiento de la señal fue una herramienta ampliamente desarrollada durante la Segunda Guerra Mundial. Desde un punto de vista histórico, el envío de mensajes cifrados, así como la recepción y análisis de los mensajes enviados por las potencias del Eje, fue pieza clave para que la guerra se decantase eventualmente por el bando Aliado. El *espectrógrafo*, inventado en 1939, era un rudimentario dispositivo eléctrico capaz de representar, en papel continuo, el espectro de Fourier de una señal a medida que éste evolucionaba en el tiempo. Los descubrimientos teóricos y prácticos acerca del tratamiento, análisis y codificación de las señales fueron secretos militares hasta el final de la guerra, de modo que el dispositivo y sus aplicaciones no fueron revelados a la comunidad científica hasta la publicación de una serie de artículos al respecto en el año 1946. En estos artículos se deduce que la transformada de Fourier había quedado definitivamente obsoleta, si bien no existían más que unos pocos apuntes teóricos acerca de herramientas alternativas que proporcionasen, como parecía primordial, la información frecuencial de la señal distribuida en el tiempo. Tal información presenta un carácter tridimensional: las componentes frecuenciales de una señal en un eje, su evolución temporal en otro, y la cantidad de energía instantánea que cada una de éstas posee, en el tercero. Las Transformaciones o Distribuciones Tiempo–Frecuencia (TFD) evolucionaron con rapidez a partir de ese momento, generando toda una base de conocimiento esencial en el posterior desarrollo de la teoría de la señal. Existen multitud de TFD, entre las que se van a destacar, por su importancia histórica y su extensa utilización, la Transformada Localizada (o Corta) de Fourier (STFT), también llamada Transformada de Gabor y la Distribución de Wigner-Ville (WVD) además de la Transformada Wavelet (WT). Sin embargo existen muchas otras [10, 11], como la Transformada del Coseno, la de Hough, la Distribución de Choi-Williams y la de Margenau-Hill. Tanto la STFT como la WVD fueron desarrolladas en los años 40. El origen último de la Transformada Wavelet hay que situarlo cuatro décadas más tarde.

1.4.2. La Transformada de Fourier Localizada o de Gabor

Por primera vez introducida en el trabajo de Dennis Gabor en 1946 [67], en la STFT (Transformada Enventanada, Transformada Corta o Transformada Localizada de Fourier), el núcleo de operación es una ventana simétrica real, $g(t)$, normalizada a uno, desplazada en tiempo un factor k y modulada en frecuencia por un factor ϕ. La STFT de una señal

$x(t)$ se puede escribir [112] como:

$$STFT_x(k,\phi) = \int_{-\infty}^{+\infty} x(t)g(t-k)e^{-j\phi t}dt \tag{1.27}$$

La STFT presenta la misma resolución en los ejes temporal y frecuencial. Es el soporte de las implementaciones del Vocoder de Seguimiento de Fase (TPV) de Quatiery y McAulay [132], así como de la Síntesis por Modelado Espectral (*Spectral Modeling Synthesis*, SMS) de Xavier Serra [148]. Cada pico del modulo de la FT en una trama (frame) del análisis es un parcial de la representación. El análisis complejo efectuado a través de la Transformada de Fourier permite obtener y seguir la fase de cada parcial de una ventana de análisis a la siguiente. La pérdida de resolución temporal provoca que no se conozca la fase en cada instante de tiempo, y se hace necesario interpolar sus valores para obtener la evolución de la Frecuencia Instantánea de cada parcial. Mediante este método, la información acerca da la Frecuencia Instantánea puede ser recuperada con gran exactitud [154].

1.4.3. La Distribución de Wigner-Ville

La Distribución de Wigner-Ville (WVD), fue concebida por J. Ville [164] a partir de una forma cuadrática ensayada por E. P. Wigner en un artículo sobre termodinámica cuántica de 1932 [171]. La WVD de una señal $x(t)$ se puede escribir como:

$$WVD_x(k,\phi) = \int_{-\infty}^{+\infty} x(k+\frac{\tau}{2})x^*(k-\frac{\tau}{2})e^{-j\phi\tau}d\tau \tag{1.28}$$

La WVD es, pues, la autocorrelación de la señal analizada. La operación marcada mediante asterisco (*) es la conjugación compleja. Tal vez la más destacada característica de la WVD sea que la distribución es definida real, al ser la FT de $x(k+\tau/2)x^*(k-\tau/2)$. Pero la WVD satisface además un gran número de propiedades matemáticas deseables: se trata de una representación unitaria que conserva la energía y es covariante ante traslaciones temporales y dilataciones frecuenciales. Compatible con filtros y modulaciones, a partir de la WVD se puede recuperar tanto la frecuencia instantánea como el retardo de grupo. En concreto, la frecuencia instantánea de una señal $x(t)$ es la frecuencia media de la WVD aplicada a la señal analítica $x_{an}(t)$ [31, 164]. En otras palabras, $f_{ins}(t)$ es el primer momento de la Distribución de Wigner-Ville respecto a la frecuencia, y queda definido por la expresión:

$$f_{ins}(k) = \frac{\int_{-\infty}^{+\infty} \varphi WVD_{x_{an}}(k,\varphi)d\varphi}{\int_{-\infty}^{+\infty} WVD_{x_{an}}(k,\varphi)d\varphi} \tag{1.29}$$

donde, evidentemente, la variable k es el tiempo. Respecto al retardo de grupo, τ_x, que no va a ser tenido en cuenta a lo largo de esta disertación, es el primer momento de la WVD

respecto al tiempo, y puede escribirse como:

$$\tau_x(\varphi) = \frac{\int_{-\infty}^{+\infty} kWVD_{x_{an}}(k,\varphi)dk}{\int_{-\infty}^{+\infty} WVD_{x_{an}}(k,\varphi)dk} \tag{1.30}$$

Además satisface las llamadas *propiedades marginales*. Para una señal $x(t)$ de energía E_x, y dada su *densidad de energía* $\rho_x(t,\omega)$, se puede escribir:

$$E_x = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \rho_x(t,\omega)dtd\omega = \int_{-\infty}^{+\infty} |x(t)|^2 dt = \int_{-\infty}^{+\infty} |X(\omega)|^2 d\omega \tag{1.31}$$

donde $|x(t)|^2$ y $|X(\omega)|^2$ son las densidades de energía en tiempo y frecuencia, respectivamente. En tal caso, las propiedades marginales se definen en las siguientes ecuaciones:

$$\int_{-\infty}^{+\infty} \rho_x(t,\omega)d\omega = |x(t)|^2 \tag{1.32}$$

y:

$$\int_{-\infty}^{+\infty} \rho_x(t,\omega)dt = |X(\omega)|^2 \tag{1.33}$$

lo que significa que, integrando la densidad de energía tiempo-frecuencia a lo largo de una de éstas variables, se obtiene la densidad de energía correspondiente a la otra [7].

Evaluada numéricamente utilizando algoritmos FFT [31], no genera dispersión temporal ni frecuencial para una delta de Dirac o sinusoide pura. Pero cuando la señal a analizar posee dos o más componentes, la WVD ofrece resultados no nulos en localizaciones inesperadas del semiplano tiempo–frecuencia. La aplicación de esta transformada queda pues limitada por estos términos de interferencia, que pueden ser parcialmente eliminados sin más que suavizar la transformada WVD_x mediante cierto núcleo θ [82], pasando a llamarse Distribución Pseudo Wigner-Ville (PWVD). En la PWVD (véase Capítulo 5, Sección 5.3.1.2), la resolución temporal y frecuencial pasan a depender de la apertura del núcleo utilizado en cada punto del plano, $\theta(k,\phi)$ [82, 112]. La eliminación de términos de interferencia por este método conlleva que algunas de las citadas propiedades matemáticas dejen de ser satisfechas, por lo que se buscará un equilibrio entre ambas características.

La obtención de la frecuencia instantánea y del retardo de grupo de la señal analizada a través de las Ecuaciones 1.29 y 1.30 no es por lo tanto fácil ni directa. Este hecho, unido al tiempo de computación que lleva obtenerla [6] y a los términos de interferencia espuria con los que se ha de convivir, son las principales limitaciones que presenta el uso de esta transformada. Esto coloca *a priori* a la WVD en desventaja con respecto a la STFT y en particular la Transformada Wavelet en la cual, como se demostrará a lo largo de esta Tesis, ninguna de estas limitaciones está realmente presente. De hecho, la Transformada Wavelet

es óptima para el estudio de señales que varían con el tiempo [123].

1.5. La Transformada Wavelet

A mediados de los 80, Grossmann, Morlet y otros [72, 73, 74, 75] introdujeron la Transformada Wavelet (WT), en la cual se calcula la convolución entre la señal a analizar y una serie de versiones desplazadas en el tiempo y comprimidas/dilatadas en frecuencia de cierta función llamada *filtro atómico* o *wavelet madre*. La WT lleva a cabo el análisis bajo filtros pasobanda de ancho de banda relativo (factor de calidad Q) constante [51].

1.5.1. Conceptos básicos de la Transformada Wavelet

Considérese la señal genérica de la Ecuación (1.1), que se repite a continuación:

$$x(t) = A(t)\cos[\Phi(t)] \tag{1.34}$$

La Transformada Wavelet Continua de una señal tal puede definirse por la expresión [50]:

$$W_x(a,b) = \int_{-\infty}^{+\infty} x(t)\Psi_{a,b}^*(t)dt = \langle x, \Psi_{a,b}\rangle = x \star \overline{\Psi}_a(b) \tag{1.35}$$

siendo * de nuevo la conjugación compleja, y $\Psi_{a,b}(t)$ la wavelet madre, wavelet de análisis o wavelet atómica, escalada en frecuencia por el factor a y desplazada en el tiempo por b:

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}}\Psi\left(\frac{t-b}{a}\right) \tag{1.36}$$

Dependiendo de si $\Psi_{a,b}(t)$ es una función real o compleja, la Ecuación (1.35) resulta ser a la vez la definición de la Transformada Wavelet Continua (CWT) y de la Transformada Wavelet Continua y Compleja (CCWT) [50]. En esta expresión, el operador $\star$ es el producto de convolución entre la señal analizada y un filtro cuya respuesta al impulso es $\overline{\Psi}_a(t)$, donde:

$$\overline{\Psi}_a(t) = \frac{1}{\sqrt{a}}\Psi^*\left(\frac{-t}{a}\right) \tag{1.37}$$

La FT de $\overline{\Psi}_a(t)$ es:

$$\widehat{\overline{\Psi}}_a(\omega) = \sqrt{a}\widehat{\Psi}^*(a\omega) \tag{1.38}$$

Por lo tanto, la Transformada Wavelet Continua de una señal es equivalente a un filtrado pasobanda de la misma. La convolución computa la WT mediante filtros pasobanda dilatados. Para más información acerca de las propiedades matemáticas generales de la CWT (CCWT), véase [50].

1.5.1.1. Admisibilidad wavelet

En la Ecuación (1.35), no todas las funciones $\Psi_{a,b}(t)$ sirven como wavelet madre. Existe la llamada *condición de admisibilidad*, que asegura un adecuado comportamiento del banco de filtros obtenido por dilatación/compresión y desplazamiento de la wavelet atómica. Tal condición de admisibilidad puede escribirse como [50]:

$$C_\psi = \int_0^{+\infty} \frac{|\hat{\psi}(\xi)|^2}{\xi} d\xi < +\infty \tag{1.39}$$

La Ecuación (1.39) asegura [39, 50, 111] que los coeficientes wavelet de la Ecuación (1.35) están bien definidos (son finitos). Si ψ es integrable al menos una vez en el sentido de Lebesgue, es decir, si $\psi \in L^1(\mathbb{R})$, y si se cumple:

$$\int_{-\infty}^{+\infty} (1 + |t|)|\psi(t)|dt < +\infty \tag{1.40}$$

En tal caso, aplicando el Teorema de Riemann-Lebesgue, $\hat{\psi}$ es continuamente diferenciable [50] y la Ecuación (1.39) sólo puede ser cierta si:

$$\psi(0) = 0 \tag{1.41}$$

Lo cual es equivalente a solicitar que la wavelet madre tenga un valor medio nulo [50, 111], es decir:

$$\int_{-\infty}^{+\infty} \hat{\psi}(x)dx = 0 \tag{1.42}$$

Se puede demostrar que las condiciones de admisibilidad de las Ecuaciones 1.39 y 1.42 son equivalentes en la práctica [50]. En el caso de wavelets complejas, estas condiciones se reflejan en que *la FT de una wavelet madre analítica debe ser real, y nula para frecuencias negativas* [111]. Es decir, una wavelet madre compleja *debe ser analítica*.

1.5.1.2. Wavelets madre

La literatura provee de una larga e interesante relación de wavelets madre, cada una de las cuales presenta diferentes características frecuenciales y temporales. Algunas de estas son:

- Haar.
- Sombrero mejicano.
- Meyer.

- Spline cuadrática.
- Spline diádica.
- Daubechies.
- Bi-ortogonal.
- Gaussiana.
- Morlet.

Cada una de estas funciones tiene una definición matemática propia, si bien no se entrará en detalles particulares sobre ninguna de ellas, remitiendo para el particular a los espléndidos trabajos de Daubechies [50], Mallat [111, 112] y el original de Landau, Pollak y Slepian [157]. Algunas de las wavelet madre citadas pueden tener soporte real o complejo, y cada una de ellas posee características particulares que pueden hacerla óptima para determinadas aplicaciones.

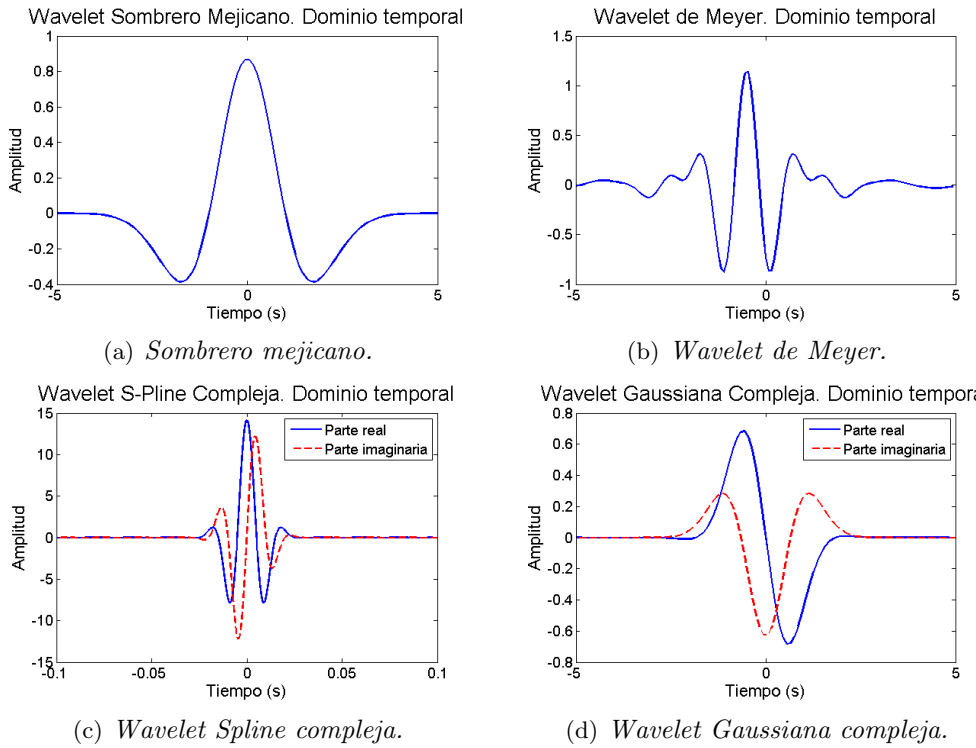
No obstante, en la Figura 1.2 se han representado cuatro de éstas wavelets madre (dos reales y dos complejas), en el dominio del tiempo, generadas con la Wavelet Toolbox de Matlab®.

En el presente trabajo, la wavelet madre empleada ha sido la Wavelet de Morlet, presentada en la Sección 1.5.2.1 y detallada en la Sección 2.2.

1.5.1.3. Resolución temporal y frecuencial

La diferencia que más evidente resulta de comparar la STFT con la WT concierne a la resolución del análisis. Como se ha afirmado en la Sección 1.4.2, la STFT presenta una resolución constante, tanto en el eje frecuencial como en el temporal. Sin embargo, de cara a analizar estructuras de la señal de tamaños muy diferentes, se haría necesario emplear funciones atómicas con diferentes soportes temporales.

Esto significa que, empleando la STFT para analizar adecuadamente eventos de diferente duración (por ejemplo, transitorios frente a sinusoides), se necesita emplear ventanas de diferentes tamaños (debido a la Relación de Incertidumbre de Heisemberg, ventanas más cortas en tiempo son más anchas en frecuencia y viceversa, lo que define las llamadas *cajas de Heisemberg*, con un ancho frecuencial ligado al ancho temporal de la misma). Sin embargo, a diferencia de la base empleada en el análisis de Fourier, el banco de filtros que se genera en el análisis wavelet lleva implícito el empleo de estas estructuras atómicas de diferentes tamaños en tiempo y frecuencia, ya que descompone la señal sobre wavelets dilatadas en frecuencia y desplazadas en tiempo [112].



(a) *Sombrero mejicano.* (b) *Wavelet de Meyer.*

(c) *Wavelet Spline compleja.* (d) *Wavelet Gaussiana compleja.*

Figura 1.2: *Diferentes wavelet madre: (a) Sombrero mejicano y (b) Wavelet de Meyer, definidas reales. Las wavelets complejas (c) Spline, orden 10 y (d) Gaussiana. Estas últimas tienen su equivalente real.*

En la Figura 1.3 se han representado en rojo y azul sendas cajas de Heisemberg correspondientes a dos wavelets centradas en frecuencias distintas (y por lo tanto de anchos de banda diferentes).

Dado que emplea un banco de filtros de Q constante, la CCWT proporciona mayor resolución frecuencial a medida que la frecuencia central de los filtros decrece. Como se puede apreciar en la figura, la resolución es variable dependiendo de la posición espectral del filtro, y la resolución temporal presenta el comportamiento complementario. Un estudio detallado de la relación entre el filtrado pasobanda de una señal y la WT puede encontrarse en [111].

Al igual que la STFT, la WT es capaz de resolver la evolución temporal de los transitorios frecuenciales. Esto requiere el uso de una wavelet analítica compleja la cual, como se ha dicho, puede separar la amplitud y fase de las componentes detectadas. La CCWT de una señal $x(t)$ depende sólo de su parte analítica, y conserva la energía para señales reales (de variable real) [112].

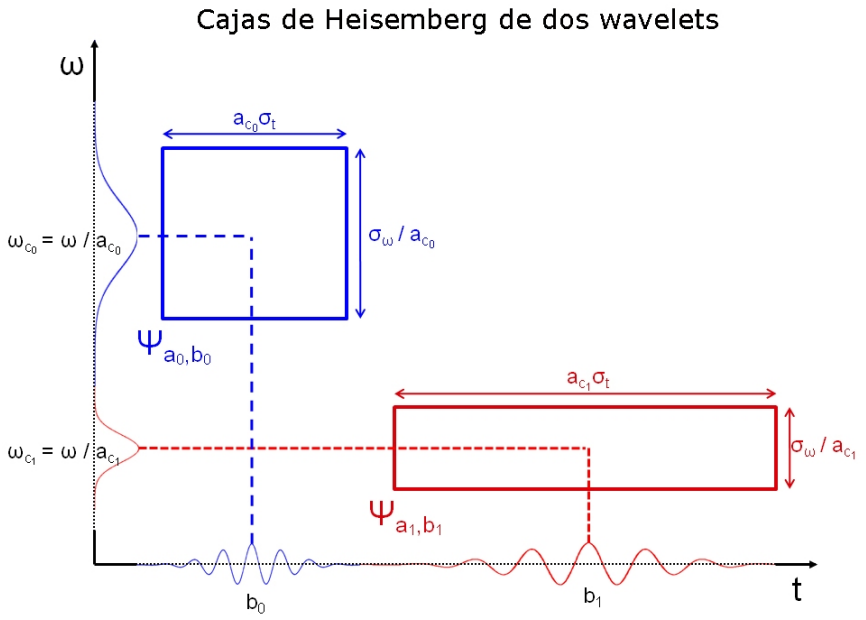


Figura 1.3: *Cajas de Heisemberg de dos wavelets. Las escalas pequeñas hacen decrecer el ancho temporal pero incrementan el soporte frecuencial, que se desplaza a frecuencias más altas. Figura inspirada en [112].*

1.5.2. La Transformada Wavelet Continua y Compleja

Al igual que la STFT, la WT puede medir la evolución temporal de transitorios frecuenciales. Esto requiere el uso de una wavelet analítica (compleja), de la cual separar la amplitud y la fase de las componentes detectadas [112]. Como se ha dicho, la Transformada Wavelet Continua y Compleja (CCWT) puede igualmente calcularse a través de la expresión dada en la Ecuación (1.35), utilizando una $\Psi_{a,b}(t)$ compleja. En tal caso, los coeficientes wavelet obtenidos pueden efectivamente ser analizados en módulo y fase [50, 73]:

$$\|W_x(a,b)\| = \sqrt{\Re[W_x(a,b)]^2 + \Im[W_x(a,b)]^2} \tag{1.43}$$

$$\Phi_x(a,b) = arg\left(\Re[W_x(a,b)] + j\Im[W_x(a,b)]\right) \tag{1.44}$$

Es posible obtener la amplitud instantánea de la señal genérica de la Ecuación (1.34), $A(t)$, a partir de la Ecuación (1.43) y su fase instantánea $\Phi(t)$ de la Ecuación (1.44). La frecuencia instantánea de $x(t)$ en la escala a_0 se obtiene a partir de la derivada temporal de (1.44), [31]:

$$f_{ins}(a_0,b) = \frac{1}{2\pi}\frac{d[\Phi_x(a_0,b)]}{db} \tag{1.45}$$

cuyo paralelismo con la Ecuación (1.9) resulta evidente.

Hay una serie de relaciones interesantes que afectan a la CCWT, la FT y la HT. Como se detallará en la Sección 1.5.2.2, a través de tales relaciones, es posible calcular los coeficientes wavelet evitando la Ecuación (1.35), en concreto a través de la inversa de la FT de la multiplicación de la FT de la señal de entrada y la FT del banco de filtros.

1.5.2.1. La Wavelet de Morlet

Como se verá más adelante, el presente trabajo se ha desarrollado a partir de un germen inicial creado en el Grupo de Audio Digital de la Universidad de Zaragoza [17], el cual presentaba una serie de limitaciones prácticas. En los primeros estadios del presente trabajo, se realizó un estudio teórico de diferentes funciones atómicas, intentando dilucidar si tales problemas experimentales de la transformada podían ser debidos a una incorrecta elección de la base. Sin embargo, la mayoría de los estudios sobre la WT coinciden en que, tomando ciertas precauciones (véase Capítulo 2, Sección 2.2.1), la Wavelet de Morlet[2] posee características que la hacen muy recomendable para el análisis de señales no estacionarias. La Wavelet de Morlet va a ser, como se ha adelantado, la función escogida como wavelet madre para el algoritmo desarrollado. Las razones principales para ésta decisión han sido básicamente tres:

- Fue la base escogida para el algoritmo inicial.
- Es resoluble analíticamente (posee soporte infinito).
- Por su FT parece óptima para la generación de un banco de filtros pasobanda (con un alto grado de no-ortogonalidad, lo cual supondrá a la postre una ventaja primordial).

Como mera introducción a la Wavelet de Morlet, se van a especificar sus definiciones matemáticas más básicas. En el dominio temporal, puede escribirse como:

$$\psi(t) = Ce^{\frac{-t^2}{2}}e^{j\xi_0 t} \tag{1.46}$$

En cuanto al dominio de la frecuencia, la transformada de Fourier de la Wavelet de Morlet es:

$$\hat{\psi}(\omega) = C'e^{-\sigma^2\frac{(\omega-\omega_0)^2}{2}} \tag{1.47}$$

En la Figura 1.4 se han representado gráficamente las Ecuaciones (1.46) y (1.47) para $C = C' = 1$ y $\omega_0 = 5$rad/s.

[2] *Jean Morlet (13 de Enero de 1931 – 27 de Abril de 2007) fue un geofísico francés que realizó trabajos pioneros en el campo del análisis wavelet en colaboración con Alex Grossman. Morlet inventó el término "wavelet"(ondelette) para describir las ecuaciones similares a las que habían existido desde la década de 1930.*

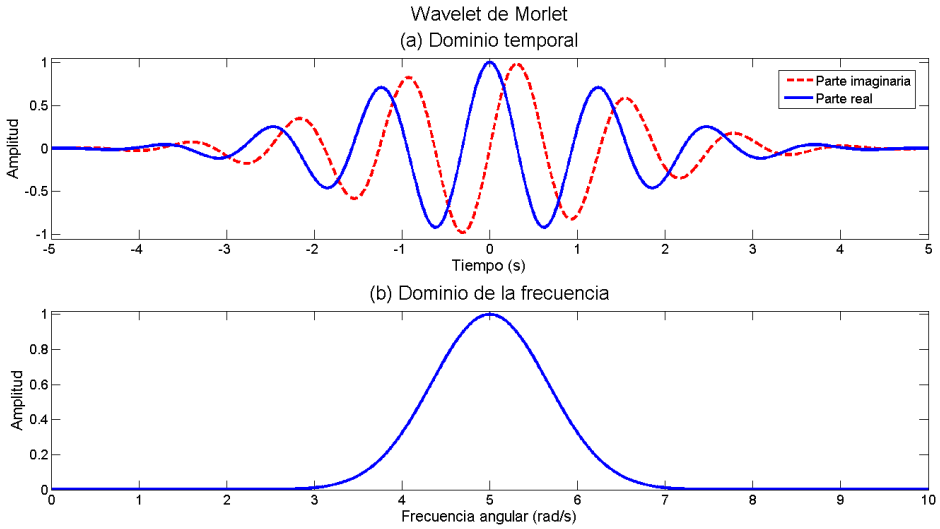


Figura 1.4: *Wavelet de Morlet. (a) Dominio temporal. La línea continua se corresponde con la parte real de la función, el trazo discontinuo con la parte imaginaria. (b) Dominio frecuencial. Gaussiana de ancho de banda σ y frecuencia central ω_0.*

Un estudio más detallado acerca de la Wavelet de Morlet finalmente utilizada y sus características principales ocupará la Sección 2.2.

1.5.2.2. Relaciones con la Transformada de Hilbert

En esta sección se van a mostrar algunas relaciones que involucran a la Transformada de Hilbert (HT) y a la CCWT, incluyendo la obtención de la señal analítica $x_{an}(t)$ a partir de la Transformada de Fourier de la señal original $x(t)$. La idea es demostrar además que el filtrado de la señal analítica a través del banco de filtros definido por $\Psi(t)$ es equivalente al filtrado de la señal original utilizando un filtro real $g(t)$ relacionado con $\Psi(t)$ de forma paralela a la relación existente entre $x(t)$ y $x_{an}(t)$.

La HT de $x(t)$ se define como:

$$HT[x(t)] = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{x(t)}{t-\tau}d\tau = x(t) \star \frac{1}{\pi t} \tag{1.48}$$

Aplicando la Trasformada de Fourier (FT) a ambos lados de la ecuación anterior, se tiene:

$$\widehat{HT}_x(\omega) = \widehat{x}(\omega)\widehat{r}(\omega) \tag{1.49}$$

donde:

$$\widehat{r}(\omega) = \widehat{\left(\frac{1}{\pi t}\right)}(\omega) \tag{1.50}$$

En esta ecuación, la FT de $1/\pi t$ es:

$$\widehat{r}(\omega) = -jsign(\omega) \tag{1.51}$$

donde $sign(\omega)$ es una función que puede tomar los valores:

$$sign(\omega) = \begin{cases} +1 & \text{si } \omega > 0 \\ 0 & \text{si } \omega = 0 \\ -1 & \text{si } \omega < 0 \end{cases} \tag{1.52}$$

Estas ecuaciones proporcionan un método para construir una señal partiendo de $x(t)$ cuyo espectro estaría formado únicamente por las frecuencias positivas del espectro de $x(t)$. A esta señal es a la que se ha llamado anteriormente señal *analítica* (ver Sección 1.2), $x_{an}(t)$. Es el momento de detallar la obtención de $x_{an}(t)$. Partiendo de la expresión general de una onda $x(t)$, dada en las Ecuaciones (1.1) y (1.34), la señal analítica de $x(t)$ se obtiene a través de su HT. Por otro lado, siempre que $x(t)$ sea asintótica, es decir, si cumple con la Ecuación (1.2), la HT es un operador que, como se ha dicho, introduce un desfase de 90° en la función a la que afecta. Sea $X(t)$ la HT de $x(t)$. Sumando:

$$x(t) + jX(t) \approx \; A(t)\cos[\phi(t)] + jA(t)\sin[\phi(t)] = A_x(t)e^{j\phi_x(t)} = x_{an}(t) \tag{1.53}$$

Como se adelantó en la Sección 1.2 este resultado se le conoce en la literatura como el *Teorema de Bedrosian*, mientras que al par de funciones ($A_x \geq 0$, $\phi_x \epsilon [0, 2\pi]$) se les conoce como el *par canónico* de $x(t)$. Utilizando la Ecuación (1.51), la FT de la señal analítica $x_{an}(t)$ viene dada por:

$$\widehat{x}_{an}(\omega) = \begin{cases} +2\widehat{x}(\omega) & \text{si } \omega > 0 \\ \widehat{x}(0) & \text{si } \omega = 0 \\ 0 & \text{si } \omega < 0 \end{cases} \tag{1.54}$$

Por otro lado, sea $g(t)$ un filtro real de la forma:

$$g(t) = Ce^{\frac{-t^2}{2\sigma^2}} cos(\omega_0 t) \tag{1.55}$$

Es posible aplicar el mismo proceso a $g(t)$ para obtener su función analítica relativa, en este caso:

$$g(t) + jG(t) \approx Ce^{\frac{-t^2}{2\sigma^2}} e^{j\omega_0 t} = \Psi(t) \tag{1.56}$$

donde $G(t)$ es la HT de $g(t)$ y $g(t)$, debe ser asintótico, cumpliendo con la Ecuación (1.2). La

ecuación así obtenida es la expresión más comúnmente utilizada de la Wavelet de Morlet, Ecuación (1.46), la wavelet madre que se va a utilizar. En un abuso de lenguaje, se puede afirmar pues que la Wavelet de Morlet es un *filtro analítico*.

Usando la FT de una convolución, Ecuación (1.49), se tiene un método alternativo de cálculo de los coeficientes wavelet $W_x(a,b)$, usualmente dados por la Ecuación (1.35):

$$W_x(a,b) = \left(x \star \overline{\Psi}_a\right)(b) = \begin{cases} FT^{-1}[2\hat{x}(\omega)\hat{g}(\omega)] & \text{si } \omega > 0 \\ FT^{-1}[\hat{x}(0)\hat{g}(0)] & \text{si } \omega = 0 \\ 0 & \text{si } \omega < 0 \end{cases} \tag{1.57}$$

Asumiendo que el parámetro b lleva implícita la evolución temporal de la Wavelet de Morlet, es posible obtener exactamente el mismo resultado calculando los coeficientes wavelet como la convolución entre la señal analítica $x_{an}(t)$ de la Ecuación (1.53) y el filtro real $g(t)$ de la Ecuación (1.55), es decir:

$$x_{an}(t) \star g(t) = \begin{cases} FT^{-1}[2\hat{x}(\omega)\hat{g}(\omega)] & \text{si } \omega > 0 \\ FT^{-1}[\hat{x}(0)\hat{g}(0)] & \text{si } \omega = 0 \\ 0 & \text{si } \omega < 0 \end{cases} \tag{1.58}$$

o lo que es lo mismo, resulta equivalente filtrar la señal real original $x(t)$ a través del banco de filtros complejo definido por la Wavelet de Morlet que filtrar la señal analítica $x_{an}(t)$ a través del filtro real $g(t)$ relativo a $\Psi(t)$. Por lo tanto, ambas Ecuaciones (1.57) y (1.58) resultan formas paralelas y equivalentes de calcular los coeficientes wavelet de la señal original, evitando así el cálculo de la integral dada por la Ecuación (1.35).

1.5.3. Fase estacionaria: crestas y esqueletos

Desde sus primeras implementaciones en computadores [100, 101] la WT se ha revelado como una poderosa herramienta en el procesado de señales no estacionarias, y particularmente en las señales de audio. Evidentemente, tanto la implementación software como los resultados y aplicaciones del análisis dependen de la wavelet madre empleada, y no sólo por su forma particular en tiempo y/o frecuencia, sino desde un punto de vista más general, en función de si se trata de wavelets reales o complejas, como se ha avanzado anteriormente.

El modelo asintótico (exponencial) de la señal y el filtro base del análisis, tanto como sus relaciones con el par canónico presentados respectivamente en las Secciones 1.2 y 1.5.2.2 son el punto de partida para introducir el concepto de fase estacionaria, y a partir de éste, los de cresta y esqueleto.

Considérese una señal $x(t)$ analítica monocromática, así como un filtro analítico $g(t)$ (como la Wavelet de Morlet). Sean $W_x(a,b)$ sus correspondientes coeficientes wavelet, que

pueden escribirse como sigue:

$$W_x(a,b) = \int_{-\infty}^{+\infty} M_{(a,b)}(t) e^{\Phi_{(a,b)}(t)} dt \tag{1.59}$$

donde:

$$M_{(a,b)}(t) = A_s(t) A_g\left(\frac{t-b}{a}\right) \tag{1.60}$$

es el módulo compuesto, y:

$$\Phi_{(a,b)}(t) = \phi_s(t) - \phi_g\left(\frac{t-b}{a}\right) \tag{1.61}$$

es la fase total de la convolución.

Asumiendo la aproximación asintótica de la Ecuación (1.2) para señal y filtro, y mediante argumentos matemáticos generales [48, 51] se deduce que la contribución principal a los coeficientes wavelet de la Ecuación (1.59) se da en los alrededores de los puntos $t_x(a,b)$ tales que:

$$\Phi'_{(a,b)}(t_x) = 0 \tag{1.62}$$

es decir, básicamente alrededor de los puntos de fase constante, por lo que este resultado es conocido como el *argumento de fase estacionaria*, y es utilizado habitualmente en la literatura para calcular los coeficientes wavelet en regiones más delimitadas del semiplano $T-F$.

Las *crestas* de los coeficientes son conjuntos de puntos (a,b) del semiplano para los cuales se cumple que $t_s(a,b) = b$. Existe una relación biunívoca a_r que delimita cada una de estas relaciones, de modo que las crestas son, por definición, curvas $\mathcal{R}$:

$$\mathcal{R} = \left\{(a,b) \in \Omega | \quad a = a_r(b)\right\} \tag{1.63}$$

siendo Ω la región del semiplano $T-F$ en la que tiene soporte la señal $x(t)$.

Por otro lado, si la wavelet base es compleja (como es el caso de los filtros analíticos que nos ocupa), la energía de la señal se concentra alrededor de los puntos del semiplano ligados a la frecuencia instantánea de la señal [31]. Por lo tanto, en cada una de las curvas dadas por los diferentes a_r, se encuentra codificada la ley de variación frecuencial correspondiente a una de las componentes de la señal analizada [51].

La restricción de los coeficientes wavelet a la cresta $\mathcal{R}$ se conoce como el *esqueleto* de la transformada. El conocimiento de la cresta y el esqueleto de la señal en el semiplano es suficiente para caracterizar la transformada y por ende la señal $x(t)$.

En los trabajos de Carmona [41, 42] y Kronland-Martinet [76, 51] y sus respectivos grupos de investigación, se sientan las bases de diferentes métodos de estimación de las

crestas y esqueletos para señales monocomponente [51], localmente monocomponente [76] y multicomponente [41, 42], así como los diferentes grados de exactitud obtenidos y las funciones de penalización relativas a cada método propuesto. Por la similitud con el método presentado en esta tesis, cabe destacar el empleo de los puntos del semiplano donde el módulo de éstos posee un máximo local [42, 41].

En nuestro caso [17] se recurre a la Ecuación (1.57) para calcular los coeficientes wavelet en el semiplano completo (lo cual hace innecesario estimar previamente las crestas y el esqueleto de la señal). Por otro lado, el seguimiento temporal de la evolución frecuencial de las componentes detectadas coincide, como se ha visto, con el concepto de cresta, y la señal restringida a tales componentes debidamente etiquetadas y analizadas proporciona el esqueleto de la transformada, que es el que se emplea para la síntesis aditiva que genera la señal de salida. Como se verá más adelante, pese a que la coincidencia entre el algoritmo de partida y la base teórica es básicamente perfecta, los resultados obtenidos así como el mismo método inicial, presentan algunas características que no parecen encajar con lo esperado. El estudio de esta aparente contradicción es uno de los objetivos puntuales de la presente disertación. Como se verá en los próximos capítulos, tal estudio ha desembocado en una nueva definición de parcial, íntimamente ligada con el esqueleto de la señal analizada.

1.6. Usos potenciales de la Transformada Wavelet en una o varias dimensiones

Además de para el análisis y síntesis de la señal de audio, aplicación para la cual se ha optimizado el uso de la CCWT, es posible utilizar la Transformada Wavelet (considerada en general) en otros múltiples campos, con resultados heterogéneos. La característica principal que habilita a la WT para poder llevar a cabo tan diferentes aplicaciones es su posibilidad de ser utilizada en múltiples dimensiones [155].

De entre todas las posibilidades, se van a destacar sólo algunas de las grandes líneas básicas de investigación internacional: la investigación en astrofísica, el procesado de señales biomédicas y de señales geofísicas. Como denominador común de todas ellas, destacaremos el procesamiento de imágenes. En cualquier caso cabe señalar que, incluso mediante una búsqueda meramente superficial, es posible encontrar ejemplos de aplicaciones en muchas otras ramas de la ciencia tan diversas como el análisis meteorológico [162], monitorización de máquinas en previsión de posibles averías [6, 36, 130] o incluso en análisis económico, donde cabe destacar los estudios de Ramsey [133, 134] y Crowley [49].

1.6.1. Astronomía y astrofísica

La WT es una de las herramientas más utilizadas en astronomía y astrofísica de los últimos diez años. En general, se ha utilizado ampliamente en múltiples aplicaciones, que van desde el filtrado y deconvolución de datos de estrellas a la detección de galaxias o la eliminación de los rayos cósmicos. En una búsqueda rápida dentro del Sistema de Datos Astrofísicos de la NASA aparecen más de 1000 artículos de investigación que contienen la palabra "wavelet" en ese período de tiempo [160]. En astrofísica, el análisis de eventos temporales y la búsqueda de periodicidades a pequeña y gran escala (milisegundos en púlsares, días en estrellas variables, años en la actividad solar, períodos de miles a millones de años en evolución estelar) es un campo de cultivo muy importante para que proliferen las aplicaciones de las representaciones tiempo–frecuencia en general.

En [80], la wavelet de Morlet es empleada en el análisis de datos heliosismológicos para demostrar que las oscilaciones solares (las cuales dependen de la estructura interna del Sol) tienen carácter no estacionario, mientras que la gran resolución temporal de los resultados obtenidos permite concluir que los modos son excitados súbitamente, en tiempos aparentemente aleatorios. El análisis wavelet de ciclos de actividad solar [62] a través de grupos de manchas (fenómeno cuasi-cíclico, con un período aproximado de 22 años, resultado del análisis de datos recogidos pacientemente por centenares de astrónomos desde el siglo XVII) refleja un segundo período subyacente con una periodicidad aproximada de un siglo. Por otro lado se demuestra que algunos eventos históricos en la actividad solar son desviaciones de la actividad normal.

El uso de wavelets no ortogonales [106] en el estudio de estructuras de nubes de gas interestelar (aplicado concretamente a nubes de CO presentes en Banard 5) refleja la estructura jerárquica de la organización del gas, posibilitando claves de su proceso de fragmentación.

Por supuesto, el tratamiento de imágenes es particularmente importante dentro de este campo de la física, ya sea en la eliminación de información no deseada o en la detección de patrones [160]. Las imágenes de un mismo cuerpo astronómico proporcionadas por telescopios situados en satélites no son idénticas unas de otras; presentan ligeras diferencias debidas a vibraciones mecánicas o al cambio en la posición orbital del satélite. Además, su resolución está limitada por parámetros de construcción (apertura de la CCD) y por el propio medio difuso interestelar. El uso de wavelets, mediante análisis comparativo de varias fotografías del mismo objeto, permite la localización y filtrado de sectores ruidosos dentro de las imágenes, y así proporciona un medio de aumentar enormemente la resolución de las mismas [172]. En astrofísica de altas energías, las imágenes tienden a presentar ruido de Poisson, el cual puede ser detectado y filtrado mediante el análisis wavelet, concretamente empleando la wavelet de Haar [94, 161].

Existe software específico, como el programa *PixInsight*[3], diseñado específicamente para los exigentes requerimientos tanto de la astrofotografía como otros campos de procesamiento de imágenes, que incluye, entre otras técnicas, varias basadas en el procesado wavelet. En la Figura 1.5 se representa un ejemplo de procesamiento, realizado con el citado programa sobre una imagen de la Nebulosa de Orión[4], *M42*. En la imagen, el procesado wavelet consigue una importante mejora en los detalles de la estructura nebular a gran escala.

(a) *Imagen original.* (b) *Imagen procesada.*

Figura 1.5: *Ejemplo de procesado de imagen astronómica empleando el programa* PixInsight*: (a) Imagen original de la nebulosa de Orión (M42). (b) Imagen procesada de la misma nebulosa, empleando una herramienta basada en la Transformada Wavelet. Los detalles aparecen mucho más definidos.*

1.6.2. Biomedicina

La transformada wavelet ha sido utilizada del mismo modo en, principalmente, dos ramas biomédicas: por un lado el análisis, caracterización y clasificación de señales biomédicas, y por otro el tratamiento de imágenes.

Respecto al tratamiento de la señal, cabe destacar el uso de la WT en el análisis de señales electrocardiográficas (ECG), electroencefalográficas (ECG), músculo-nerviosas o electromiográficas (EMG), sonidos clínicos, patrones respiratorios, tendencias de presión

[3]PixInsight *es un producto de* Pleiades Astrophoto S. L.*, una compañía de desarrollo de software que goza de un equipo internacional con experiencia en la astronomía, la fotografía, las matemáticas y la ingeniería de software. Para más información, véase* `http://www.pixinsight.com/`.

[4]*Ejemplos de procesado con* PixInsight *empleando el llamado algoritmo de Transformada Wavelet de Alto Rango Dinámico* (High Dynamic Range Wavelet Transform)*, desarrollado por Vicent Peris (astrofotógrafo del Observatorio Astronómico de la Universidad de Valencia y miembro de PTeam),* `http://www.pixinsight.com/examples/HDRWT/examples/en.html`. *Publicado con permiso de los autores.*

arterial y secuencias de ADN. En [4] se discute el empleo de la WT y otras herramientas de análisis para el tratamiento de las señales biomédicas. El análisis wavelet produce una descomposición tiempo–frecuencia que es capaz de separar las componentes individuales de la señal de un modo más efectivo que la STFT [2].

La búsqueda bibliográfica llevada a cabo se ha centrado principalmente en el estudio de la señal ECG, por motivos que se aclararán en la Sección 1.6.5. Las diferentes zonas de interés de un ECG (segmentos PR y ST, intervalos PR y QT, complejo QRS) aparecen marcadas en la Figura 1.6[5].

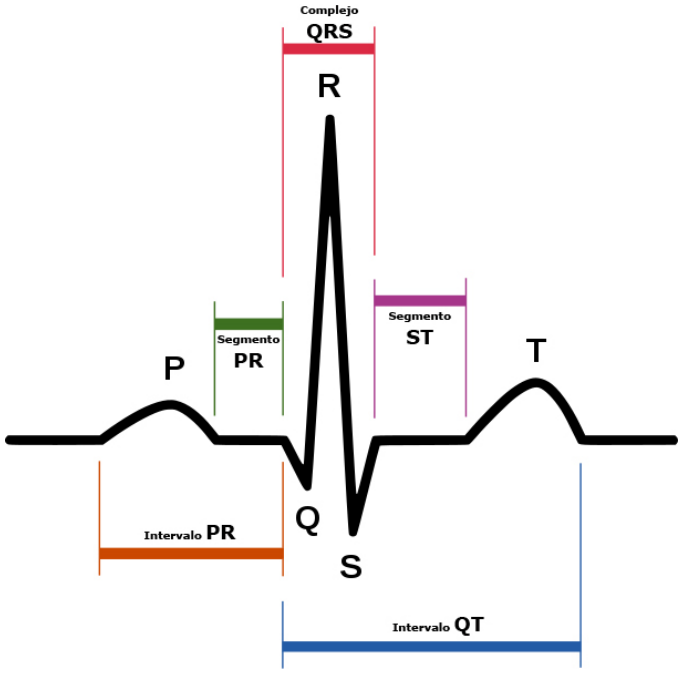


Figura 1.6: *Puntos característicos de una onda electrocardiográfica: onda P, complejo QRS, onda T.*

La correcta detección de los tiempos de entrada-salida de cada punto característico de la señal ECG (complejo QRS, ondas P y T, Figura 1.6) [2, 109], es vital para la clasificación automática de potenciales cardiopatías. En [145] se emplean la transformada wavelet real en base diádica, la DWT y la wavelet compleja para la detección del complejo QRS. Las anomalías en el segmento ST del ECG pueden ser síntomas de isquemia, necrosis e infarto. La wavelet de Morlet ha sido empleada para detectar este segmento, así como el complejo QRS [95]. Una vez detectados los puntos característicos del ECG se puede emplear la WT junto con modelos ocultos de Markov para la caracterización de la señal [141].

[5] *Figura original procedente de* wikipedia [12]*, traducción al castellano del autor.*

La señal electroencefalográfica, ECG, se emplea clínicamente para investigar desórdenes cerebrales. En [81] se emplea la WT unida a una red neuronal para analizar y clasificar señales ECG procedentes de cerebros normales, pacientes de esquizofrenia y de trastorno obsesivo-compulsivo.

En cuanto al tratamiento de imágenes particularizado a esta rama de la ciencia, por ejemplo en [83] se emplea la Transformada Wavelet Compleja de Árbol Dual (Dual Tree Complex Wavelet Transform, DT-CWT) para el filtrado y la reducción de ruido.

1.6.3. Geofísica

En cuanto a la geofísica, la importancia potencial de cualquier avance en el desarrollo de herramientas de análisis en tiempo y frecuencia son más que evidentes, habida cuenta de la magnitud que con desgraciada frecuencia alcanzan los eventos naturales sísmicos y vulcanológicos.

Las potenciales aplicaciones geofísicas de la WT en sus múltiples formas (wavelet continuas reales, discretas, ortogonales y paquetes wavelet) son enormes [103]: Análisis de procesos no estacionarios que contienen características de múltiples escalas, detección de singularidades, análisis de fenómenos transitorios, procesos fractales y multifractales. La compresión de la señal basada en wavelets (campo en principio no aplicable a la versión de la transformada aquí presentada) ha sido explotada para el estudio de diferentes procesos, incluyendo la precipitación del espacio-tiempo, la medida remota de flujos hidrológicos, análisis de turbulencias atmosféricas, de la cobertura del dosel oceánico, estudios de topografía de la superficie de la tierra, la batimetría del fondo marino, y la clasificación de las olas del mar causadas por viento. En [44] se emplea la CWT para resolver las limitaciones de la STFT en el análisis tiempo–frecuencia de señales sismográficas, mientras que en [102] la misma transformada se emplea para analizar, filtrar y modelar la propagación de ondas s y p, así como para estimar, a través de los datos del semiplano T–F, la velocidad de fase y de grupo en los datos de un sismograma. En [176] se emplea la Transformada Wavelet Compleja Empaquetada (Complex Wavelet Packet Transform, CWPT) para el estudio de señales sísmicas de fase no lineal. En [118] se demuestran las ventajas de la DT-CWT para para reducir el ruido y los *artefactos* de procesamiento en el análisis de señales sísmicas a partir de datos acústicos.

En vulcanología [142], la CCWT ha multiplicado la profundidad de los análisis de movimiento magmático subterráneo, demostrando la relación entre transitorios anómalos de auto-potenciales eléctricos y fracturas superficiales del terreno durante la fase eruptiva. En [14] se aprovecha la coherencia del análisis wavelet para interpretar las señales recogidas en un reconocido cráter, clasificándolas en temblores volcánicos de fondo, transitorios causados por las frecuentes explosiones internas y ruido sísmico indefinido.

En la actualidad se está trasladando el uso del algoritmo CWAS propuesto en esta Tesis

a la detección precisa de los tiempos de llegada de las ondas transversales y superficiales (S, P, Love y Rayleigh, véase Figura 1.7) en terremotos lejanos y cercanos, combinando un filtrado de alta precisión con análisis de modos de polarización, localización de ondas refractadas y de tiempos de onset.

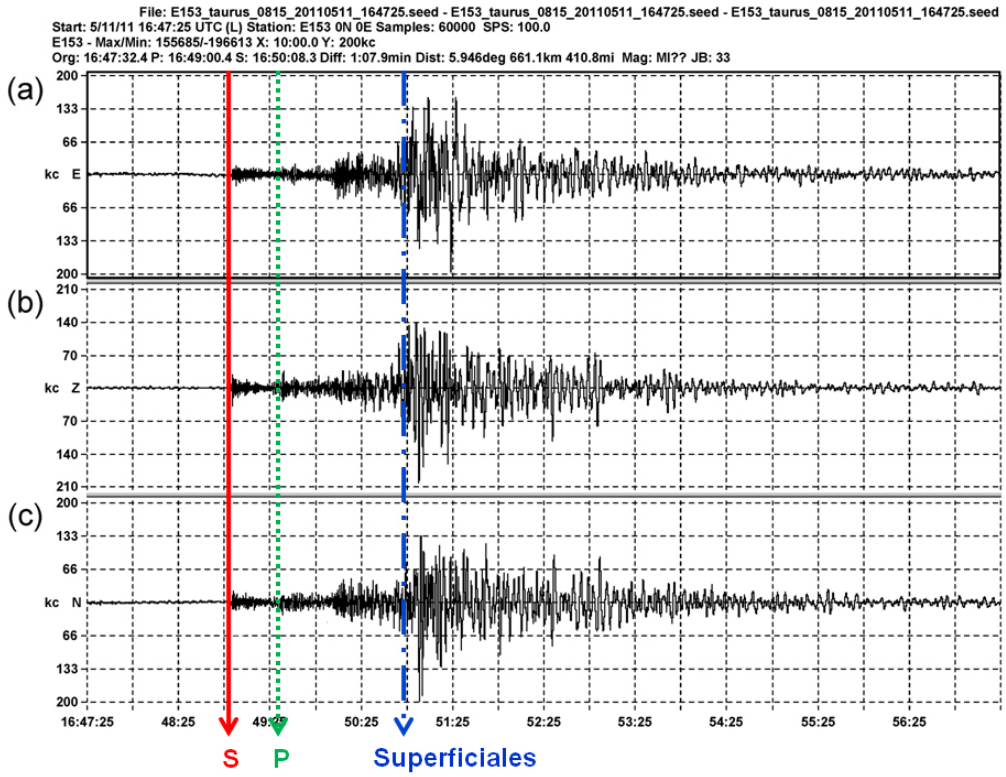


Figura 1.7: *Sismograma correspondiente al terremoto de Lorca (11-05-2011). Los instantes de llegada de las ondas S, P y superficiales (Love-Rayleigh) están marcados en rojo, verde y azul. (a) Componente E-W. (b) Componente vertical. (c) Componente N-S.*

1.6.4. Procesado de imagen

La Transformada Wavelet presentada puede extenderse con facilidad a dos o más dimensiones [50], ampliándose de este modo el abanico de sus posibles aplicaciones. Las exigencias matemáticas para una wavelet definida en más de una dimensión es una extensión de las que se han presentado en la Sección 1.5.1. Por motivos de extensión, no se va a entrar en detalles acerca de éstas definiciones y propiedades, para lo cual se recomienda consultar el citado trabajo de Daubechies [50].

Dentro de las aplicaciones de la WT en dos dimensiones, el procesado de imagen es una de las más destacadas. El empleo de filtros logarítmicos está muy extendido en el análisis de imágenes por computador, por ejemplo en la detección de contornos (uno de los trabajos de investigación del autor, previo requisito a la obtención del DEA versó sobre este tema). Acerca del empleo de la WT en este tipo de aplicaciones, se destacará el trabajo de José Ramón Beltrán, en cuya Tesis doctoral y publicaciones relacionadas [25, 26, 27, 28, 29] se trata del análisis y clasificación de límites y contornos en imágenes empleando la wavelet de Mallat y Zhong [113].

Por su facilidad para la descomposición de la señal en diferentes bandas de frecuencia, el análisis wavelet resulta especialmente adecuado par el procesamiento de imágenes [68] ya que:

- Se puede mejorar el factor de compresión de datos, dado que las imágenes tienden a presentar un espectro más uniforme, con la mayor parte de la energía concentrada en la parte de baja frecuencia.

- Al igual que, como se verá, ciertos modelos wavelet encajan con el comportamiento del sistema auditivo humano, modelos bidimensionales encajan con el comportamiento de nuestro sistema de visión. Esto permite reducir la presencia del ruido de alta frecuencia, permitiendo el ajuste paramétrico del nivel de distorsión producido por la compresión.

- Se reduce el impacto de compresión por bloques (como en los archivos jpeg), al analizarse la imagen en bloque.

- En las imágenes naturales pueden coexistir zonas muy homogéneas con cambios bruscos (contornos, esquinas, sombras). Al permitir el análisis de secciones estacionarias y transitorios con buenos resultados, el análisis wavelet resulta de nuevo una herramienta muy adecuada en estos casos.

1.6.5. Conclusiones

Las citadas aplicaciones de la WT son el resultado de una búsqueda no demasiado exhaustiva de información bibliográfica. Con esto no se pretende otra cosa que demostrar que las potencialidades de la transformada van mucho más allá de la rama del audio, donde se va a centrar el presente trabajo. En [138] aparece un interesante resumen de las propiedades principales del análisis wavelet en el procesado de señal, así como una visión resumida de sus posibles aplicaciones en varios campos. Para una revisión más exhaustiva de las diferentes aplicaciones de la Transformada Wavelet en cualquiera de sus formas, véase [155].

Como conclusión, cabe destacar que la precisión y coherencia de los resultados que ofrece la herramienta que se va a presentar en los próximos capítulos, permiten ser optimistas

respecto su posible utilización en el abanico de aplicaciones citado, si bien es evidente que las potencialidades de la técnica no aseguran su aplicabilidad a problemas muy distantes del audio. Por descontado, el algoritmo CWAS podría no ser apropiado para resolver algunos de los problemas que se han propuesto, y podría ser ineficiente o no presentar ventajas evidentes respecto a técnicas ya existentes, en otros.

Por otro lado, muchos de los cambios necesarios para la adaptación del algoritmo CWAS a otras ramas de la ciencia no resultarían triviales; sin ir más lejos, el análisis wavelet propuesto en esta disertación es un análisis en una dimensión, mientras que, en aplicaciones como el procesado de imágenes (ya sea genérico o aplicado a la medicina, a microscopía, cristalografía, etc.) se trata de una técnica bidimensional. El análisis sismográfico o en el de señales biomédicas como las procedentes de EEG y ECG, por el contrario, en principio tan sólo requeriría de una adaptación del banco de filtros a las necesidades del análisis. De hecho, en este momento se está ultimando la primera aproximación del algoritmo CWAS al análisis de señales electrocardiográficas [86]. Sin embargo, en el mejor de los casos estas aplicaciones futuribles se encuentran todavía en sus primeros estadios.

1.7. Objetivos y estructura de la disertación

Como se ha adelantado, basándose en las ideas expuestas en los trabajos de los grupos de Richard Kronland-Martinet y René Carmona y sus respectivos colaboradores, los doctores José Ramón y Fernando Beltrán del Departamento de Ingeniería Electrónica y Comunicaciones de la Universidad de Zaragoza desarrollaron un primer prototipo algorítmico basado en la CCWT en el año 2003 [17]. Sin embargo, esta primera versión del algoritmo presentaba dos serias limitaciones: la primera de ellas, el proceso de resíntesis requería una renormalización a la energía de la señal de cara a obtener resultados coherentes. En segundo lugar, ante señales cuya frecuencia instantánea variase saltando entre diferentes bandas de análisis, la señal reconstruida presentaba un evidente (y audible) rizado, difícil de encajar con la teoría. Es aquí donde el autor de la presente disertación se incorpora al Grupo de Audio Digital, y bajo los auspicios del Doctor José Ramón Beltrán, se inicia el presente trabajo de investigación.

1.7.1. Objetivos generales y específicos

El objetivo general de esta Tesis Doctoral es *demostrar que la Transformada Wavelet Continua y Compleja puede convertirse en una herramienta muy eficiente en la extracción de características de alto nivel de la señal de audio.*

En cuanto a los objetivos específicos iniciales, son tres:

1. *Encontrar el origen de los problemas del algoritmo original y, si es posible, superarlos.*

2. *Desarrollar diferentes aplicaciones de la CCWT, explorando sus posibilidades prácticas.*

3. *Realizar un estudio comparativo entre los algoritmos desarrollados y sus equivalentes empleando otras TFD, siempre que esta comparación sea factible.*

Como se verá a lo largo de los siguientes cinco capítulos, tanto el objetivo general como los específicos han sido adecuadamente cubiertos, superándose en muchos casos las expectativas iniciales.

1.7.2. Estructura del trabajo

Esta disertación contiene un total de seis capítulos básicos, el primero de los cuales (el que nos ocupa en este momento) es un estudio general del estado del arte, así como una introducción a los conceptos más elementales que serán empleados en adelante.

En el Capítulo 2 se introducen las soluciones halladas a las mencionadas limitaciones (Objetivo 1). Estas soluciones han desembocado en el llamado algoritmo de Síntesis Aditiva por Wavelets Complejas (CWAS), detallado en el Capítulo 3, donde se detallan sus bloques constitutivos más importantes. A continuación, en los Capítulos 4 y 5 se detallan las aplicaciones principales que se han ido desarrollando durante el período investigador que cubre la presente disertación (Objetivo 2), las cuales se pueden dividir en dos categorías: la extracción de características generales de alto nivel de las señales de audio y una aplicación específica más concreta, la separación ciega de notas musicales en señales monaurales. En el Capítulo 5 se presenta además una comparativa entre el algoritmo CWAS y otras herramientas de análisis de sonido como SMS, las rutinas espectrográficas de alta resolución o la Time-Frequency Toolbox de Matlab® (Objetivo 3).

Para cerrar el estudio, las conclusiones principales del mismo aparecen en el Capítulo 6.

Como ampliación de los resultados teóricos y experimentales, se han incluido un total de seis anexos, la mayoría de los cuales está directamente ligado a un capítulo concreto de esta Tesis.

Capítulo 2

Consideraciones matemáticas sobre la Transformada Wavelet Continua y Compleja

Índice

"Todo saber tiene de ciencia
lo que tiene de matemática".
Jules Henri Poincaré (1854-1912).
Matemático, físico y filósofo francés.

En este capítulo se detallará el modelo matemático subyacente al algoritmo propuesto, que ha servido como fuente de inspiración para las diferentes mejoras que se han ido introduciendo en la técnica de análisis y caracterización de señales acústicas, y que han derivado en el estado final defendido en esta disertación. Se encontrará el origen de las discrepancias entre la teoría y la práctica de la Transformada Wavelet Continua y Compleja, y se propondrán soluciones a estos problemas, cubriéndose de este modo el primero de los objetivos de esta Tesis.

2.1. Introducción

Como se ha dicho al final del capítulo anterior, entre la teoría (fundamentalmente los trabajos de los grupos de Richard Kronland-Martinet [51, 72, 76, 100, 101] y René A. Carmona [41, 42]) y la práctica (la primera aproximación a un algoritmo funcional basado en la CCWT por parte de los doctores José Ramón y Fernando Beltrán), mediaba una serie de dificultades prácticas que requerían, muy probablemente, un enfoque distinto, en mayor o menor medida, al propuesto en el algoritmo de partida [17]. Para desentrañar el origen de los problemas encontrados (recapitulando de nuevo, la renormalización forzada y la aparición de evidentes rizados incluso en los parciales de variación frecuencial no excesiva) parecía absolutamente necesario un análisis matemático riguroso, pero a la vez más accesible que el propuesto por Kronland y Carmona, y sus respectivos equipos de investigadores.

Las primeras aportaciones del autor fueron precisamente en este campo [19, 20]. Se trata de analizar, de forma coherente y matemáticamente sólida [22], las características de la CCWT del modo más general pero a la vez más práctico posible. Esto pasa por resolver explícitamente las ecuaciones de los coeficientes wavelet de la Ecuación (1.35), analizándolos a continuación exclusivamente en módulo. En efecto, al igual que otras distribuciones Tiempo–Frecuencia, en la CCWT el espectrograma wavelet de fase contiene información temporal muy interesante, pero será omitido por completo en adelante (excepto deducciones puntuales que se detallarán en alguna de las secciones siguientes), puesto que la información

modular es suficiente en general para caracterizar la señal en bandas, y con ésta caracterización y los coeficientes complejos originales, los parciales constituyentes detectados se pueden obtener fácilmente, como se verá en el Capítulo 3.

Lamentablemente, sólo en ciertos casos muy concretos, los coeficientes wavelet resultan calculables de forma exacta. Como se verá a continuación, se hace necesario revisar la wavelet madre utilizada en nuestro análisis, añadiéndole ciertos parámetros de control que permiten la generación de bancos de filtros versátiles de forma muy simple. A continuación, se procede a analizar la señal más sencilla (tono puro de amplitud constante) bajo el prisma de la CCWT. Después se aumenta progresivamente la dificultad del análisis, intentando sacar las conclusiones más generales del mismo.

2.2. La Wavelet de Morlet revisada

Como se adelantó en el Capítulo 1, la wavelet madre escogida para el análisis presentado en esta disertación es la wavelet de Morlet.

Aparte de las razones esgrimidas en el capítulo anterior y más allá de la gran redundancia contenida en los coeficientes wavelet obtenidos a través de esta función, la wavelet de Morlet ha sido escogida como función atómica puesto que admite insertar en ella sutiles cambios de fondo que permiten flexibilizar enormemente la estructura del banco de filtros que se obtiene a partir de su base.

2.2.1. La Wavelet de Morlet básica

La wavelet de Morlet, en su definición inicial más completa, puede escribirse matemáticamente [50, 101] como:

$$\psi(y) = Ce^{\frac{-y^2}{2}} \left(e^{j\xi_0 y} - e^{\frac{\xi_0^2}{2}} \right) \tag{2.1}$$

Estrictamente hablando, esta función no satisface las condiciones formales para una wavelet madre, presentadas en la Sección 1.5.1.1. Sin embargo, si ξ_0 es suficientemente grande, y por lo tanto su transformada de Fourier $\hat{\Psi}(\xi)$ se anula para $\xi < 0$, las correcciones necesarias son numéricamente despreciables. Como demuestra Daubechies [50], es suficiente con tomar:

$$\xi_0 \geq 5 \tag{2.2}$$

2.2.2. Cambios sobre la Wavelet de Morlet

De cara a obtener un control más riguroso del banco de filtros, hemos introducido un parámetro de ancho de banda σ en la Ecuación (2.1):

$$y = \frac{t}{\sigma} \tag{2.3}$$

Por lo tanto, en la Ecuación (2.1):

$$\xi_0 = \sigma\omega_0 \tag{2.4}$$

y de esta forma, la expresión conduce a:

$$\psi(t) = Ce^{-\frac{t^2}{2\sigma^2}}\left(e^{j\omega_0 t} - e^{-\frac{\sigma^2\omega_0^2}{2}}\right) \approx \; Ce^{-\frac{t^2}{2\sigma^2}}e^{j\omega_0 t} \tag{2.5}$$

En el dominio del tiempo, la wavelet de Morlet es una exponencial compleja modulada por una gaussiana de anchura $2\sqrt{2}/\sigma$. La transformada de Fourier de la wavelet de Morlet compleja dilatada es:

$$\hat{\psi_a}(\omega) = C'e^{-\sigma^2\frac{(a\omega-\omega_0)^2}{2}} \tag{2.6}$$

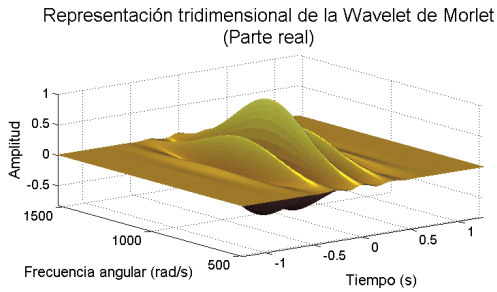
La ecuación anterior representa una gaussiana centrada en la frecuencia ω_0/a. C y C' son las constantes de normalización de la función atómica en los dominios temporal y frecuencial, respectivamente. Tomando esta función base, la CCWT es equivalente un banco de filtros pasobanda de respuesta en frecuencia dada por la Ecuación (2.6).

Las Ecuaciones (2.5) y (2.6) han sido presentadas en la Figura 1.4, para la cual se ha tomado $C = C' = 1$, $a = 1$ y $\omega_0 = 5$rad/s. En la Figura 2.1 aparece una representación tridimensional de la superficie que representa la wavelet de Morlet en el semiplano tiempo–frecuencia (frecuencia central, $\omega_c = 1000$rad/s).

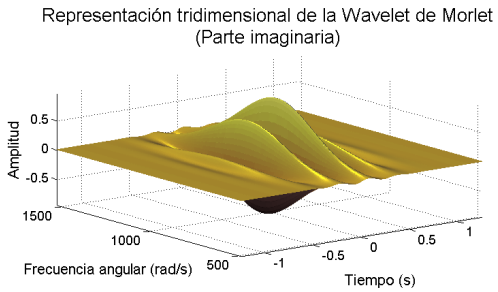
Concretamente, en la Figura 2.1(a) se ha representado la parte real de ésta wavelet madre, en la Figura 2.1(b) su parte imaginaria, y para terminar, en la Figura 2.1(c) la combinación de ambas en el valor absoluto (módulo). Esta última figura resulta especialmente ilustrativa, puesto que muestra que el módulo de la Wavelet de Morlet es en realidad una gaussiana doble, en tiempo y frecuencia.

Como se avanzó en el capítulo anterior, cada uno de los miembros del banco de filtros utilizado en el análisis debe ser asintótico según se define en [48, 53]. Utilizando las expresiones para T y B dadas por las Ecuaciones (1.6) y (1.7) respectivamente [31, 67, 156], la aproximación asintótica, Ecuación (1.8) [157], se verifica si:

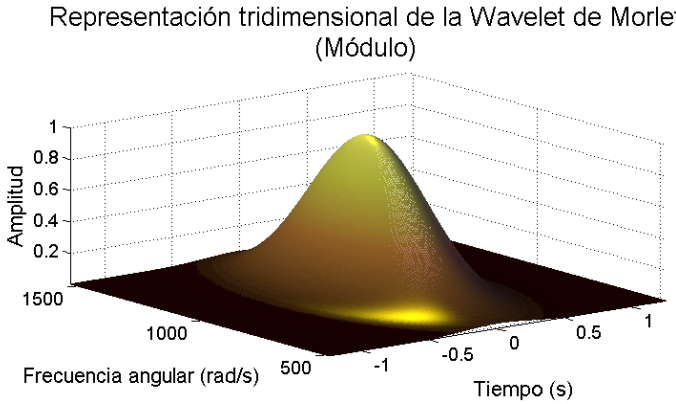
$$\sigma^2\omega_0^2 \geq 25 \tag{2.7}$$



(a) *Parte real.*



(b) *Parte imaginaria.*



(c) *Valor absoluto.*

Figura 2.1: *Representación 3D de la Wavelet de Morlet en el semiplano tiempo–frecuencia: (a) Parte real. (b) Parte imaginaria. Ambas, adecuadamente combinadas, generan (c) Módulo de la Wavelet de Morlet.*

La relación entre este resultado y la Ecuación (2.2) resulta evidente.

2.3. Metodología (I): Análisis

Como se ha especificado en varias ocasiones, el algoritmo de partida presenta básicamente dos limitaciones: por un lado, la aparente dependencia de la señal de entrada en el proceso de renormalización, algo que, en principio, no encaja con las predicciones de la literatura. En segundo lugar, la aparición de un rizado muy evidente en señales cuya frecuencia instantánea evoluciona entre bandas adyacentes (rizado tanto más evidente cuantos más de estos cambios de banda existan). Para encontrar el origen del error y de este modo poder subsanarlo, se requería un estudio matemático riguroso de los coeficientes wavelet. Desgraciadamente, la literatura presenta evidentes dificultades prácticas, ya que se trata de estudios de enorme complejidad matemática, muy difícilmente visionables y/o programables

en su estado original.

El primer paso consiste en estudiar los coeficientes wavelet desde un punto de vista menos general, pero que a la vez permita la obtención explícita y exacta de tales coeficientes, con la intención de encontrar en tal proceso de visualización las claves que permitan superar las evidentes limitaciones de la técnica inicial.

A continuación, se procede a la resolución expresa de los coeficientes wavelet para el caso más simple: una señal pura monocromática de amplitud constante.

2.3.1. Señal de AM monocromática y de amplitud constante

Como se ha avanzado, se trata de calcular de forma exacta los coeficientes wavelet de la Ecuación integral (1.35), es decir:

$$W_x(a,b) = \int_{-\infty}^{+\infty} x(t)\Psi^*_{a,b}(t)dt \tag{2.8}$$

siendo (recuérdese la Sección 1.5.1.2):

$$\Psi(t) = Ce^{-\frac{t^2}{2\sigma^2}}e^{j\omega_0 t} \tag{2.9}$$

expresión que debe asimismo cumplir con la Ecuación (1.36). En este análisis se conserva además una constante de normalización genérica, es decir:

$$\Psi_{a,b}(t) = C\ \Psi\left(\frac{t-b}{a}\right) \tag{2.10}$$

Este filtro-base va a ser sustituido en la Ecuación (2.8), para el caso de una señal de la forma:

$$x(t) = A_1\cos(\omega_1 t + \varphi) \tag{2.11}$$

donde φ representa la fase inicial de la señal.

Dadas las características de los coeficientes wavelet implícitas en la Ecuación (2.8), es decir, soporte infinito e integrabilidad de Lebesgue al menos a nivel 1, es evidente que no todas las señales de entrada resultan resolubles de manera explícita siguiendo directamente esta expresión. Sin embargo, las conclusiones de los análisis más simples pueden aplicarse intuitivamente a casos más complicados. Se va a resolver de forma detallada este caso, mientras que en el resto de los casos se explicitarán únicamente los pasos más importantes, extrayendo las correspondientes conclusiones.

La expresión de partida es:

$$W_x(a,b) = \int_{-\infty}^{+\infty} A_1\cos(\omega_1 t + \varphi)Ce^{-\frac{(t-b)^2}{2a^2\sigma^2}}e^{-j\omega_0\frac{t-b}{a}}dt \tag{2.12}$$

Dado que:

$$\cos\chi = \frac{e^{j\chi} + e^{-j\chi}}{2} \tag{2.13}$$

es evidente que se deben resolver las dos integrales siguientes:

$$I_1 = \frac{A_1 C}{2} \int_{-\infty}^{+\infty} e^{j(\omega_1 t+\varphi)} e^{-\frac{(t-b)^2}{2a^2\sigma^2}} e^{-j\omega_0 \frac{t-b}{a}} dt \tag{2.14}$$

$$I_2 = \frac{A_1 C}{2} \int_{-\infty}^{+\infty} e^{-j(\omega_1 t+\phi)} e^{-\frac{(t-b)^2}{2a^2\sigma^2}} e^{-j\omega_0 \frac{t-b}{a}} dt \tag{2.15}$$

Ambas integrales son del tipo *Gaussiano*. La solución de este tipo de ecuaciones es:

$$\int_{-\infty}^{+\infty} e^{-(\alpha t^2+\beta t+\gamma)} dt = \sqrt{\frac{\pi}{\alpha}} e^{(\frac{\beta^2-4\alpha\gamma}{4\alpha})} \tag{2.16}$$

Reordenando términos en I_1, se tiene:

$$I_1 = \frac{A_1 C}{2} e^{j(\frac{\omega_0 b}{a}+\varphi)} \int_{-\infty}^{+\infty} e^{-\frac{t^2-[2b-j2\sigma^2 a(a\omega_1-\omega_0)]t+b^2}{2\sigma^2 a^2}} dt \tag{2.17}$$

Aplicando la Ecuación (2.16) a la integral I_1 de (2.17), se obtiene:

$$I_1 = \frac{A_1 C}{2} \sqrt{2\pi\sigma^2 a^2} e^{j\varphi} e^{-\frac{a^2\sigma^2\omega_1^2}{2}} e^{-\frac{\sigma^2\omega_0^2}{2}} e^{a\omega_1\omega_0\sigma^2} e^{j\omega_1 b} \tag{2.18}$$

para lo cual, se ha tomado simplemente:

$$\alpha = \frac{1}{2\sigma^2 a^2} \tag{2.19}$$

$$\beta = \frac{-2b + j2\sigma^2 a(a\omega_1 - \omega_0)}{2\sigma^2 a^2} \tag{2.20}$$

$$\gamma = \frac{b^2}{2\sigma^2 a^2} \tag{2.21}$$

En cuanto a I_2, procediendo de manera similar:

$$I_2 = \frac{A_1 C}{2} \sqrt{2\pi\sigma^2 a^2} e^{-j\varphi} e^{-\frac{a^2\sigma^2\omega_1^2}{2}} e^{-\frac{\sigma^2\omega_0^2}{2}} e^{-a\omega_1\omega_0\sigma^2} e^{-j\omega_1 b} \tag{2.22}$$

Sumando ambas integrales, y sabiendo que:

$$\cosh\chi = \frac{e^{\chi} + e^{-\chi}}{2} \qquad y \qquad \sinh\chi = \frac{e^{\chi} - e^{-\chi}}{2} \tag{2.23}$$

relación de la que se obtiene, finalmente:

$$W_x(a,b) = \frac{A_1 C}{2}\sqrt{2\pi\sigma^2 a^2}e^{-\frac{\sigma^2}{2}(\omega_0^2+a^2\omega_1^2)} \\ \Big[\cosh(a\omega_1\omega_0\sigma^2)\cos(\omega_1 b+\varphi) + j\sinh(a\omega_1\omega_0\sigma^2)\sin(\omega_1 b+\varphi)\Big] \tag{2.24}$$

En esta expresión, se tomará como constante de normalización:

$$C = \sqrt{\frac{2}{\pi}}\frac{1}{a\sigma} \tag{2.25}$$

lo cual conduce a resultados interesantes, como se verá a continuación.

La Ecuación (2.24) puede ser analizada en términos de módulo y fase, a través de las Ecuaciones (1.43) y (1.44) respectivamente. Con el valor de C anteriormente mencionado, resultan ser:

$$\|W_x(a,b)\|^2 = 2A_1^2 e^{-\sigma^2(\omega_0^2+a^2\omega_1^2)}\Big\{\cosh(2a\omega_1\omega_0\sigma^2) + \cos\big[2(\omega_1 b+\varphi)\big]\Big\} \tag{2.26}$$

y:

$$\Phi[W_x(a,b)] = \arctan\big[\tanh(\sigma^2\omega_0\omega_1 a)\tan(\omega_1 b+\varphi)\big] \tag{2.27}$$

El máximo del módulo de la Ecuación (2.26) se da para la escala $a = a_1 = \omega_0/\omega_1$. En tal caso, resulta:

$$-\sigma^2(\omega_0^2 + a^2\omega_1^2)|_{a=a_1} = -2\sigma^2\omega_0^2 \tag{2.28}$$

y

$$2a\omega_0\omega_1\sigma^2|_{a=a_1} = 2\sigma^2\omega_0^2 \tag{2.29}$$

Y por lo tanto, la Ecuación (2.26) se reduce a:

$$\begin{aligned}\|W_x(a_1,b)\|^2 &= 2A_1^2 e^{-2\sigma^2\omega_0^2}\Big\{\cosh(2\sigma^2\omega_0^2) + \cos\big[2(\omega_1 b+\varphi)\big]\Big\} \\ &= 2A_1^2 e^{-2\sigma^2\omega_0^2}\Big\{\frac{e^{2\sigma^2\omega_0^2}+e^{-2\sigma^2\omega_0^2}}{2} + \cos\big[2(\omega_1 b+\varphi)\big]\Big\}\end{aligned} \tag{2.30}$$

Si $\sigma^2\omega_0^2$ es suficientemente grande, $e^{-2\sigma^2\omega_0^2} \to 0$ y por lo tanto:

$$\begin{aligned}\|W_x(a_1,b)\|^2 &\approx 2A_1^2 e^{-2\sigma^2\omega_0^2}\Big\{\frac{e^{2\sigma^2\omega_0^2}}{2} + \cos\big[2(\omega_1 b+\varphi)\big]\Big\} \\ &\approx 2A_1^2 e^{-2\sigma^2\omega_0^2}\frac{e^{2\sigma^2\omega_0^2}}{2} = A_1^2\end{aligned} \tag{2.31}$$

en conclusión:

$$\|W_x(a,b)\| \approx A_1 \tag{2.32}$$

Por otro lado, teniendo en cuenta que en la escala $a_1 = \omega_0/\omega_1$, siempre que $\chi = \sigma^2\omega_0^2$ sea lo suficientemente grande, se tiene:

$$\tanh\chi = \frac{\sinh\chi}{\cosh\chi} = \frac{e^{\chi} - e^{-\chi}}{e^{\chi} + e^{-\chi}} \approx \frac{e^{\chi}}{e^{\chi}} = 1 \tag{2.33}$$

la Ecuación (2.27) de la fase, evaluada en la escala a_1 resulta:

$$\Phi[W_x(a_1,b)] \approx \arctan\left[\tan(\omega_1 b + \varphi)\right] = \omega_1 b + \varphi \tag{2.34}$$

Dicho de otro modo, el módulo de los coeficientes wavelet encierra toda la información necesaria para reconstruir la señal: por un lado, su máximo está localizado en un factor de escala directamente relacionado con la fase instantánea de la señal. Por otro lado, evaluado sobre ese factor de escala y convenientemente renormalizado, el módulo coincide con la amplitud (instantánea) de la señal.

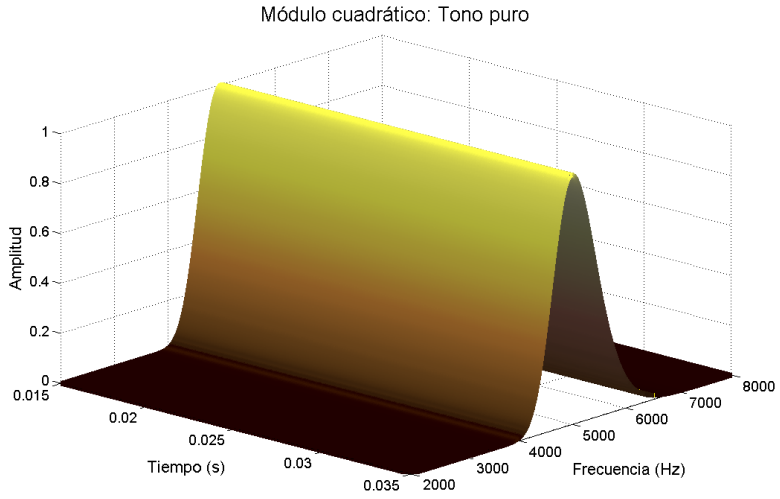


Figura 2.2: *Módulo cuadrático de los coeficientes wavelet para una señal de amplitud constante* $A_1 = 1$ *y frecuencia constante* $f_1 = 5kHz$.

En la Figura 2.2 se ha representado la superficie definida por la Ecuación (2.26) para una señal de amplitud unidad y frecuencia pura $f_1 = 5kHz$. Se puede observar cómo el máximo de los coeficientes wavelet se da en la frecuencia adecuada (o dicho de otro modo,

en la escala adecuada $a_1 = \omega_0/\omega_1$). Para tal factor de escala, obsérvese cómo el valor de la amplitud del módulo coincide con el valor de la amplitud de la onda estudiada.

En esta figura y en las equivalentes del resto del presente Capítulo, se ha tomado como valor de los parámetros de control $\omega_0 = 20$ y $\sigma = 0.6148$. El valor tomado para σ queda bastante lejos del utilizado en el algoritmo de análisis, pero resulta apropiado para lo que se pretende mostrar.

La relación entre el resultado obtenido y las conclusiones alcanzadas por Kronland-Martinet y Carmona y sus respectivos grupos para este ejemplo concreto son evidentes. Sin embargo, no se ha tratado explícitamente en ningún momento ni con las crestas de los coeficientes, ni con el esqueleto de la señal: estos dos términos son exactamente los que se han rescatado de forma implícita en el análisis. Por otro lado, este método de análisis tiene la ventaja respecto a los trabajos mencionados de que es posible conocer explícitamente la forma de los coeficientes wavelet en función de las características de la señal de entrada. Como inconveniente, se ha perdido la generalidad del análisis, por lo que se hace necesario resolver ejemplos más complicados antes de poder deducir consecuencias de mayor alcance.

2.3.2. Señal de AM monocromática y de amplitud variable

La siguiente pregunta a contestar es si dada una señal de amplitud instantánea variable, se obtienen las mismas conclusiones. En este caso, se trabajará con una señal del estilo:

$$x(t) = A(t)\cos(\omega_1 t) \tag{2.35}$$

donde la fase inicial φ de a la Ecuación (2.11) se ha ignorado, por simplicidad.

Dado que de nuevo las integrales obtenidas son Gaussianas, parece conveniente tomar una envolvente también Gaussiana para la señal de entrada, puesto que proporciona a la vez una variabilidad muy importante a la amplitud instantánea de la señal sin que ésta pierda por ello su soporte infinito. Por lo tanto, sea:

$$x(t) = A_1 e^{-\frac{(t-t_0)^2}{2\varrho^2}} \cos(\omega_1 t) \tag{2.36}$$

Para que esta señal sea asintótica, es necesario que su variación temporal sea mucho más lenta que su variación frecuencial, es decir, es necesario que, sea cual sea ω_1 (esto es, en todas las escalas a de medida), se cumpla:

$$\varrho^2 \gg \sigma^2 a^2 \tag{2.37}$$

Integrando los coeficientes wavelet, tras el correspondiente cálculo se llega a la expresión:

$$W_x(a,b) = \frac{A_1 C}{2} e^{-\frac{(b-t_0)^2}{2(\varrho^2+\sigma^2a^2)}} e^{-\frac{(\omega_0^2+\omega_1^2a^2)\sigma^2\varrho^2}{2(\varrho^2+\sigma^2a^2)}} \Big[\cosh\big(\frac{\sigma^2\varrho^2\omega_0\omega_1 a}{\varrho^2+\sigma^2a^2}\big) \cos\big(\frac{\varrho^2\omega_1 b}{\varrho^2+\sigma^2a^2}\big) + j \sinh\big(\frac{\sigma^2\varrho^2\omega_0\omega_1 a}{\varrho^2+\sigma^2a^2}\big) \sin\big(\frac{\varrho^2\omega_1 b}{\varrho^2+\sigma^2a^2}\big)\Big] \tag{2.38}$$

En la sección anterior se dedujo que en el módulo de los coeficientes wavelet se encuentra toda la información necesaria (pues, como se ha visto, el máximo del módulo se encuentra sobre la escala que señala unívocamente a la fase de la señal y, adecuadamente normalizado, en esa escala su valor es el de la amplitud de la señal). A continuación, se procede por lo tanto a analizar exclusivamente este módulo. Tomando C como:

$$C = \sqrt{\frac{2(\varrho^2+\sigma^2a^2)}{\pi\varrho^2\sigma^2a^2}} \approx \sqrt{\frac{2}{\pi}}\frac{1}{a\sigma} \tag{2.39}$$

y utilizando las adecuadas expresiones hiperbólicas y trigonométricas, se obtiene finalmente para el módulo cuadrático:

$$\|W_x(a,b)\|^2 = 2A_1^2 e^{-\frac{(b-t_0)^2}{\varrho^2+\sigma^2a^2}} e^{-\frac{(\omega_0^2+\omega_1^2a^2)\sigma^2\varrho^2}{\varrho^2+\sigma^2a^2}} \Big[\cosh\big(\frac{2\sigma^2\varrho^2\omega_0\omega_1 a}{\varrho^2+\sigma^2a^2}\big) + \cos\big(\frac{2\varrho^2\omega_1 b}{\varrho^2+\sigma^2a^2}\big)\Big] \tag{2.40}$$

En la Figura 2.3 se ha representado gráficamente la superficie descrita por la Ecuación (2.40). En este caso, la frecuencia de la señal es $f_1 = 5kHz$, mientras que la gaussiana temporal está centrada en $t_0 = 0.025$ segundos, siendo $A_1 = 1$ y $\varrho = 1.2\ 10^{-3}$. Obsérvese el parecido con la Figura 2.1(c).

Se puede comprobar fácilmente que el máximo de la Ecuación (2.40), en el eje de escalas, se encuentra en la escala $a = a_1 = \omega_0/\omega_1$. Para este valor de escala, simplificando las exponenciales y las funciones hiperbólicas como en el caso anterior, se tiene:

$$\|W_x(a_1,b)\|^2 \approx A_1^2 e^{-\frac{(b-t_0)^2}{\varrho^2}} \tag{2.41}$$

Siempre que se asuma la condición que permite aproximar la constante de normalización C de le Ecuación (2.39), esto es, que la gaussiana temporal es más ancha que la frecuencial, y por lo tanto que la Ecuación (2.37) es válida.

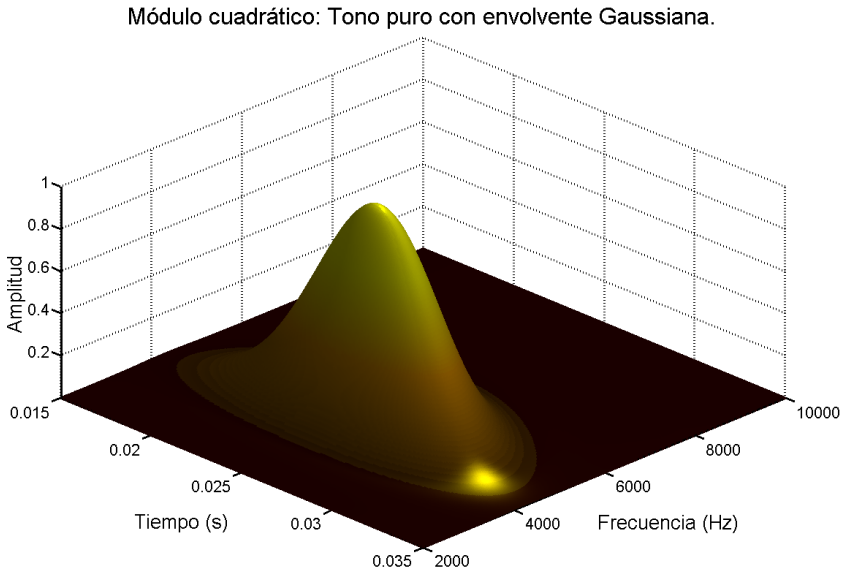


Figura 2.3: *Módulo cuadrático de los coeficientes wavelet para una señal de amplitud variable (con envolvente Gaussiana) y frecuencia constante* $f_1 = 5kHz$.

2.3.3. Señal de AM multicomponente, de amplitudes constantes

Supóngase a continuación que la señal de entrada es la suma de n sinusoides cada una de ellas de amplitud constante A_i y frecuencia angular pura ω_i, es decir:

$$x(t) = \sum_{i=1}^{n} A_i \cos(\omega_i t) \tag{2.42}$$

Por simplificar, se ha tomado de nuevo una fase inicial nula para cada componente. Caso de no ser así, los términos de fase inicial φ_i para cada componente se podrían incluir con facilidad, jugando exactamente el mismo papel que el obtenido para ellos en la Sección 2.3.1.

Sustituyendo cada coseno por su exponencial compleja, la Ecuación (2.8) se convierte en la suma de $2n$ integrales gaussianas, una pareja equivalente a I_1 e I_2 de las Ecuaciones (2.14) y (2.15) para cada una de las n componentes de la señal. Todas esas integrales resultan ser analíticamente resolubles, obteniéndose unos coeficientes wavelet complejos de la forma:

$$W_x(a,b) = \sqrt{2\pi} a \sigma C e^{-\frac{\sigma^2\omega_0^2}{2}} \sum_{i=1}^{n} \left\{ A_i e^{-\frac{\sigma^2\omega_i^2 a^2}{2}} \left[\cosh(\sigma^2\omega_0\omega_i a)\cos(\omega_i b) + \right.\right. \\ \left.\left. + j \sinh(\sigma^2\omega_0\omega_i a)\sin(\omega_i b) \right] \right\} \tag{2.43}$$

Nuevamente, la expresión anterior se puede simplificar, utilizando las aproximaciones adecuadas para las funciones hiperbólicas. Es decir, considerando las relaciones de la Ecuación (2.23) donde $\chi = \sigma^2\omega_0\omega_i a$, y teniendo en cuenta que la influencia de las exponenciales negativas es despreciable siempre que $\sigma^2\omega_0\omega_i a$ sea suficientemente grande, lo cual, como en los casos anteriores, está garantizado por la Ecuación (2.7), se sigue una nueva expresión para los coeficientes wavelet:

$$W_x(a,b) \approx \sum_{i=1}^{n} \sqrt{\frac{\pi}{2}} a\sigma C A_\alpha e^{-\frac{\sigma^2(\omega_i a-\omega_0)^2}{2}} e^{j\omega_i b} \tag{2.44}$$

Para recuperar la amplitud instantánea de cada parcial A_i de la Ecuación (2.44), se toma una constante C de valor:

$$C = \sqrt{\frac{2}{\pi}} \frac{1}{a\sigma} \tag{2.45}$$

Con la cual el módulo cuadrático de los coeficientes wavelet resulta ser:

$$\begin{aligned} \|W_x(a,b)\|^2 \approx & \sum_{i=1}^{n} A_i^2 e^{-\sigma^2(\omega_i a-\omega_0)^2} + \\ & + \sum_{i\neq k=1}^{n} A_i e^{-\frac{\sigma^2}{2}(\omega_i a-\omega_0)^2} cos(\omega_i b) \cdot A_k e^{-\frac{\sigma^2}{2}(\omega_k a-\omega_0)^2} cos(\omega_k b) \end{aligned} \tag{2.46}$$

En la Figura 2.4 aparece la Ecuación (2.46) para una señal compuesta por tres componentes de amplitudes constantes $A_1 = 0{,}5$, $A_2 = 1$ y $A_3 = 1/\sqrt{2}$ y frecuencias puras $f_1 = 3kHz$, $f_2 = 5kHz$ y $f_3 = 9kHz$. El módulo cuadrático de los coeficientes wavelet para esta señal describe la superficie representada, en la cual, localmente, cada una de las tres gaussianas mostradas se corresponde con el resultado que se hubiese obtenido según la Ecuación (2.26), representado gráficamente en la Figura 2.2, para cada una de las tres componentes. Es decir, se resuelve perfectamente la componente frecuencial de cada tono presente, así como su amplitud. Este resultado, sorprendentemente sencillo, es en realidad engañoso. La clave se encuentra en que las frecuencias de las componentes presentes están suficientemente separadas entre sí.

En efecto, en la Ecuación (2.46) podemos distinguir dos tipos de términos. En primer lugar, los que obedecen a la expresión $A_i^2 e^{-\sigma^2(\omega_i a-\omega_0)^2}$, cada uno de los cuales se corresponde con una de las diferentes componentes de la señal, considerada de forma aislada. No es difícil ver la relación de estos con la Ecuación (2.26).

Por otro lado, los términos tipo $A_i e^{-\frac{\sigma^2}{2}(\omega_i a-\omega_0)^2} cos(\omega_i b) \cdot A_k e^{-\frac{\sigma^2}{2}(\omega_k a-\omega_0)^2} cos(\omega_k b)$, que afectan a los parciales $i - simo$ y $k - simo$ de la señal, siempre que $i \neq k$, son los llamados *términos de intermodulación*. Tales términos, considerados matemáticamente, se pueden clasificar en dos categorías: en primer lugar, si las componentes i y k involucradas se en-

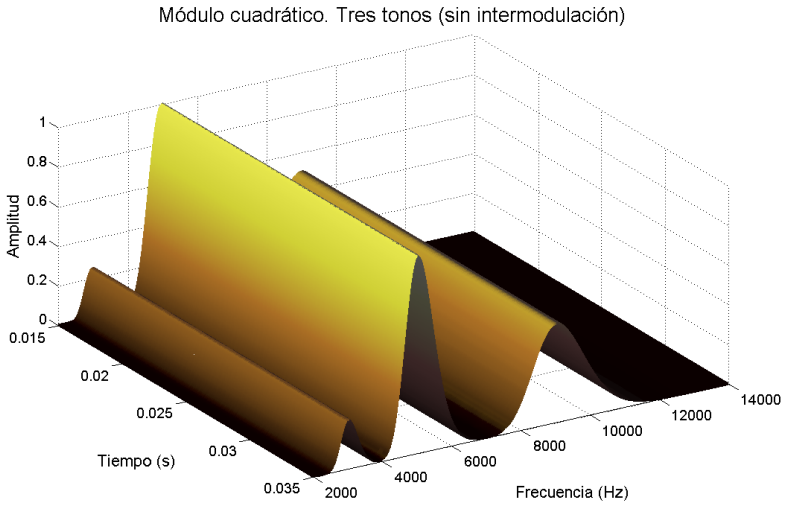


Figura 2.4: *Módulo cuadrático de los coeficientes wavelet para una señal compuesta por tres tonos de amplitudes* $A_1 = 0{,}5$, $A_2 = 1$ *y* $A_3 = 1/\sqrt{2}$ *(constantes) y frecuencias* $f_1 = 3kHz$, $f_2 = 5kHz$ *y* $f_3 = 9kHz$ *(también constantes).*

cuentran suficientemente separadas en frecuencia, entonces se verifica que $e^{-\frac{\sigma^2}{2}(\omega_i a-\omega_0)^2}$, $e^{-\frac{\sigma^2}{2}(\omega_k a-\omega_0)^2} \to 0$ y por lo tanto la intermodulación es despreciable. Sin embargo, para componentes próximas entre sí (por ejemplo, dos parciales batiendo), estos términos no sólo pueden no resultar despreciables, si no que de hecho llegan a afectar de forma clara y evidente a la localización y valor de los máximos en el semiplano $T - F$.

2.3.3.1. Intermodulación

Los términos de intermodulación aparecen reflejados en la Figura 2.5. En ella se ha representado la superficie dada por la Ecuación (2.46) para el caso de una señal compuesta por tres componentes de amplitudes $A_1 = 0.5$, $A_2 = 1/\sqrt{2}$ y $A_3 = 1$ (constantes) y frecuencias $f_1 = 3$kHz, $f_2 = 4$kHz y $f_3 = 5$kHz (también constantes). En este caso, los valores escogidos para σ y ω_0 no proporcionan suficiente resolución en el eje frecuencial. Aunque los máximos aún aparecen localizados aproximadamente en las posiciones adecuadas y alcanzan los valores esperados, se puede adivinar cierta oscilación en amplitudes y frecuencias (más evidentes en el tono de mayor energía, por su posición relativa). Entre dos vecinos cualesquiera, aparecen zonas de amplitud variable, que se corresponden con la interferencia entre los correspondientes parciales dada por la Ecuación (2.46).

Si bien las modulaciones en frecuencia se limitan al ancho de banda de los filtros impli-

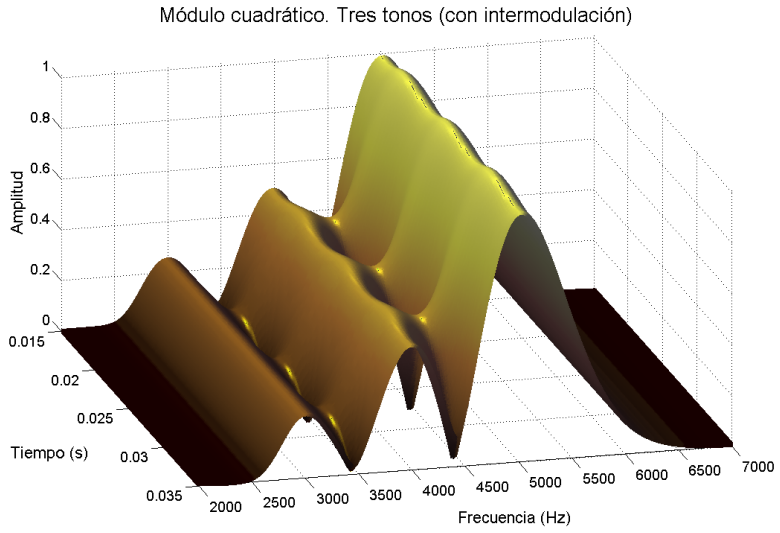


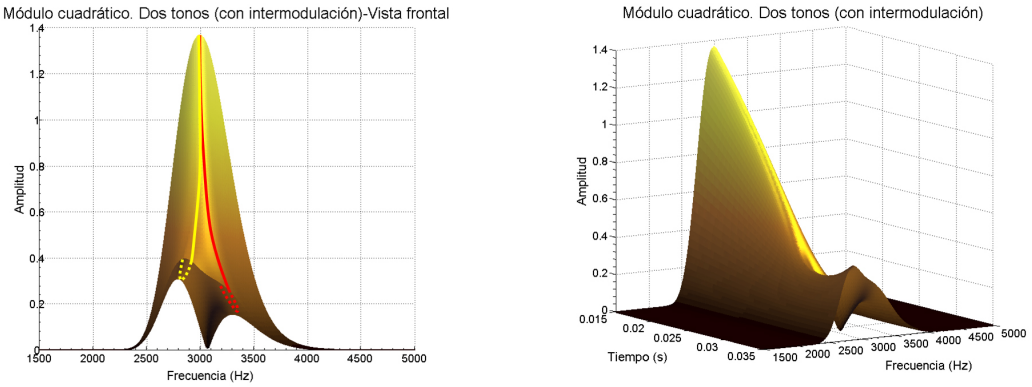
Figura 2.5: *Módulo cuadrático de los coeficientes wavelet para una señal compuesta por tres tonos próximos. La proximidad frecuencial se hace evidente en los términos de intermodulación entre pares.*

cados (y por lo tanto tienden a no ser excesivas en parciales de baja frecuencia, pero pueden alcanzar valores elevados en la zona de alta), la intermodulación, $A_i e^{-\frac{\sigma^2}{2}(\omega_i a-\omega_0)^2} cos(\omega_i b) \cdot A_k e^{-\frac{\sigma^2}{2}(\omega_k a-\omega_0)^2} cos(\omega_k b)$, es tanto más grave cuanto más parecidas entre sí sean las características de los parciales mezclados. En las Figuras 2.6(a) y 2.6(b) se presentan sendas vistas detalladas de un par de parciales de amplitudes A_1 =1 y A_2 =0.95, y frecuencias f_1 =2.95kHz y f_2 =3.05kHz. En la Figura 2.6(a) quedan reflejadas las oscilaciones en frecuencia de los máximos (cuyas trayectorias aparecen resaltadas en colores amarillo y rojo). En la vista en perspectiva de la Figura 2.6(b) es posible distinguir con claridad la magnitud de la oscilación en las amplitudes instantáneas detectadas. Estas pueden llegar a oscilar teóricamente entre los valores extremos $A_1 + A_2$ y $|A_1 - A_2|$, hecho muy importante en el desarrollo teórico/práctico detallado en el Anexo II.d.1.2.

2.3.4. Aproximación cuadrática general

Para terminar, se obtendrá de forma explícita los coeficientes wavelet para una señal asintótica general, de la forma:

$$x(t) = A(t)\cos[\phi(t)] \tag{2.47}$$



(a) *Vista frontal.* (b) *Perspectiva* 45°.

Figura 2.6: *Dos parciales muy próximos. Detalles de los términos de intermodulación: (a) Modulación en frecuencia de los máximos (trayectorias resaltadas en amarillo y rojo), (b) Modulaciones en amplitud.*

Se ha incido varias veces en que la Ecuación (2.8) no puede ser resuelta analíticamente para el caso más general. Pero tal hecho resulta soslayable, siempre que se le demande una característica nueva a la señal de entrada, de cara a hacer el cálculo aproximado que se desarrollará en esta Sección. Tal característica es que, en las proximidades de cada t_0, la frecuencia de $x(t)$ admita una expansión en serie de Taylor lineal o cuadrática [112]:

$$\phi(t) = \phi(t_0) + (t - t_0)\frac{d[\phi(t)]}{dt}\Big|_{t=t_0} + \frac{1}{2}(t - t_0)^2\frac{d^2[\phi(t)]}{dt^2}\Big|_{t=t_0} + o(t^3) \tag{2.48}$$

Se ha escogido la aproximación cuadrática por ser más general que el caso lineal.

La condición asintótica se traduce directamente a una aproximación cuasi-estática. De este modo, si la amplitud instantánea varía de forma muy lenta con respecto a la información generada en los términos exponenciales, es posible extraer *localmente* $A(t)$ de la integral dada en la Ecuación (2.8). Reemplazando además la fase $\phi(t)$ por su aproximación cuadrática evaluada en $t_0 = b$, se obtiene:

$$\begin{aligned} W_x(a,b) \approx \frac{CA(b)}{2} e^{-\frac{b^2}{2\sigma^2 a^2}} e^{j\frac{\omega_0 b}{a}} \Bigg[\int_{-\infty}^{+\infty} e^{-\left(\frac{t^2}{2\sigma^2 a^2} - \frac{tb}{\sigma^2 a^2}\right)} e^{-j\frac{\omega_0 t}{a}} e^{j\theta(b)} dt + \\ + \int_{-\infty}^{+\infty} e^{-\left(\frac{t^2}{2\sigma^2 a^2} - \frac{tb}{\sigma^2 a^2}\right)} e^{j\frac{\omega_0 t}{a}} e^{-j\theta(b)} dt \Bigg] = I_1 + I_2 \end{aligned} \tag{2.49}$$

donde:

$$\theta(b) = h_0(b) + h_1(b)t + h_2(b)t^2 + o(t^3) \tag{2.50}$$

siendo:

$$h_2(b) = \frac{\phi''(b)}{2} \tag{2.51}$$

$$h_1(b) = \phi'(b) - 2\phi''(b)b \tag{2.52}$$

$$h_0(b) = \phi(b) - \phi'(b)b + \frac{\phi''(b)}{2}b^2 \tag{2.53}$$

De cara a que la Ecuación (2.8) sea analíticamente resoluble, es necesario considerar despreciable el término $o(t^3)$ de la Ecuación (2.49). Esto se traduce en que las expresiones calculadas a partir de este momento vendrán acompañadas de una función de *penalización* tanto más despreciable cuanto mejor sea la aproximación en serie de Taylor para la fase en cada punto. Resolviendo la integral gaussiana correspondiente, los coeficientes wavelet obtenidos pueden escribirse como:

$$W_x(a,b) = [2N(a,b)]^{\frac{1}{2}} e^{-\xi(a,b)} \Big\{ \cosh[u(a,b)]\cos[\nu(a,b)] + j\sinh[u(a,b)]\sin[\nu(a,b)] \Big\} \tag{2.54}$$

Por lo que el módulo cuadrático se puede expresar:

$$\|W_x(a,b)\|^2 = N(a,b)e^{-2\xi(a,b)} \big\{ \cosh[2u(a,b)] + \cos[2\nu(a,b)] \big\} \tag{2.55}$$

y la fase queda:

$$\Phi(a,b) = \arctan\{\tanh[u(a,b)]\tan[\nu(a,b)]\} \tag{2.56}$$

En estas ecuaciones, los términos $N(a,b)$, $\xi(a,b)$ y $u(a,b)$ y $\nu(a,b)$ son:

$$N(a,b) = \frac{C^2 A_1^2}{2} \frac{2\pi\sigma^2 a^2}{\sqrt{1 + 4h_2^2\sigma^4 a^4}} \tag{2.57}$$

$$\xi(a,b) = \frac{4b^2h_2^2\sigma^2a^2 + (\omega_0^2 + a^2h_1^2)\sigma^2 + 4bh_2h_1a^2\sigma^2}{2(1 + 4h_2^2\sigma^4a^4)} \tag{2.58}$$

$$u(a,b) = \frac{\sigma^2\omega_0 a(2h_2b + h_1)}{1 + 4h_2^2\sigma^4a^4} \tag{2.59}$$

y:

$$\nu(a,b) = \frac{h_2b^2 + h_1b + h_0(1 + 4h_2^2\sigma^4a^4) - (\omega_0^2 + a^2h_1^2)\sigma^4a^2h_2}{1 + 4h_2^2\sigma^4a^4} + \frac{arctan(2h_2\sigma^2a^2)}{2} \tag{2.60}$$

En tanto en cuanto la expansión en serie de Taylor de $\phi(t)$ sea suficientemente precisa, se pueden tomar en consideración ciertas aproximaciones en todas estas expresiones. Estas aproximaciones son, básicamente, $\cosh(x) \approx \sinh(x) \approx e^x/2$ y $4^4a^4h_2^2 \ll 1$. Con tales

aproximaciones y tomando la constante de normalización C de la wavelet de Morlet como:

$$C = \sqrt{\frac{2(1+4h_2^2\sigma^4a^4)^{1/2}}{\pi\sigma^2a^2}} \approx \sqrt{\frac{2}{\pi}}\frac{1}{a\sigma} \tag{2.61}$$

se obtiene, para los coeficientes wavelet, la expresión:

$$W_x(a,b) = A(b)e^{-\frac{\sigma^2}{2}[\phi'(b)a-\omega_0]^2}e^{j\phi(b)}e^{j\varphi(a,b)} \approx A(b)e^{-\frac{\sigma^2}{2}[\phi'(b)a-\omega_0]^2}e^{j\phi(b)} \tag{2.62}$$

En esta expresión, es:

$$\varphi(a,b) = -\sigma^4a^2h_2(b)[h_1(b)a-\omega_0]^2 + \frac{\arctan[2\sigma^2a^2h_2(b)]}{2} \ll 1 \forall a,b \tag{2.63}$$

Por lo tanto:

$$\|W_x(a,b)\|^2 \approx A(b)^2e^{-\sigma^2\left[[2h_2(b)b+h_1(b)]a-\omega_0\right]^2} = A(b)^2e^{-\sigma^2[\phi'(b)a-\omega_0]^2} \tag{2.64}$$

Nótese que se ha obtenido de nuevo la expresión de una gaussiana cuya frecuencia central sigue a la frecuencia de la señal, en este caso a través del término a_b, siendo:

$$a_b = \frac{\omega_0}{2h_2(b)b+h_1(b)} = \frac{\omega_0}{\phi'(b)} = \frac{f_0}{f_{ins,x}(b)} \tag{2.65}$$

Como adelanto a futuras decisiones, a continuación se procederá a la integración de los coeficientes wavelet de la Ecuación (2.62) en el eje de escalas (a). Sea $q(b)$ el resultado. Resulta ser:

$$q(b) = \int_0^{+\infty} W_x(a,b)da \approx \frac{\sqrt{2\pi}}{\sigma\phi'(b)}A(b)e^{j\phi(b)} \tag{2.66}$$

Comparando esta expresión con la de la onda de entrada $x(t)$ de la Ecuación (2.47), es posible comprobar que lo que se acaba de obtener es, básicamente, el par canónico de la señal (en este caso sobrepesado por un factor $\sqrt{2\pi}/\sigma\phi'(b)$ el cual puede ser conocido, y por lo tanto corregido, siguiendo la frecuencia instantánea de la señal en cada instante de tiempo b). La utilidad de esta Ecuación va más allá del resultado cualitativo obtenido (muy importante, en todo caso), ya que es la base que permite superar la más importante limitación de la Transformada cuantificada (discreta), como se verá más adelante. Ahora bien, el proceso de integración en una variable continua es equivalente al sumatorio en el caso de una variable discreta. Por la tanto en la Ecuación (2.66) se encuentra la clave que permite obtener el parámetro de sobrepeso λ presentado en la Sección 3.6 del próximo capítulo.

Como ejemplo característico del resultado obtenido, se puede estudiar el caso de un *chirp lineal*. Un chip lineal es una señal de audio que obedece a la Ecuación (2.47) para la cual,

la fase instantánea es de la forma:

$$\phi(t) = \alpha t^2 + \beta t + \gamma \tag{2.67}$$

Es evidente que en este caso la aproximación en serie de Taylor de la fase coincide con la misma fase, y por lo tanto no habremos de añadir ningún término de penalización. Los coeficientes wavelet obedecen a las expresiones calculadas anteriormente, y definen una superficie en el semiplano $T-F$ que ha sido representada en la Figura 2.7 para una señal de amplitud constante $A(t) = A_1 = 1$ y parámetros α =62500, β =5000 y γ =20. La frecuencia instantánea de tal señal es de 5000Hz en $t = 0$, 6875Hz en $t = 0.015$ y 9375Hz en $t = 0.035$. En la figura se pueden comprobar visualmente tales valores (de forma aproximada). Obsérvese cómo la frecuencia instantánea es la recta que une dos cualesquiera de estos puntos.

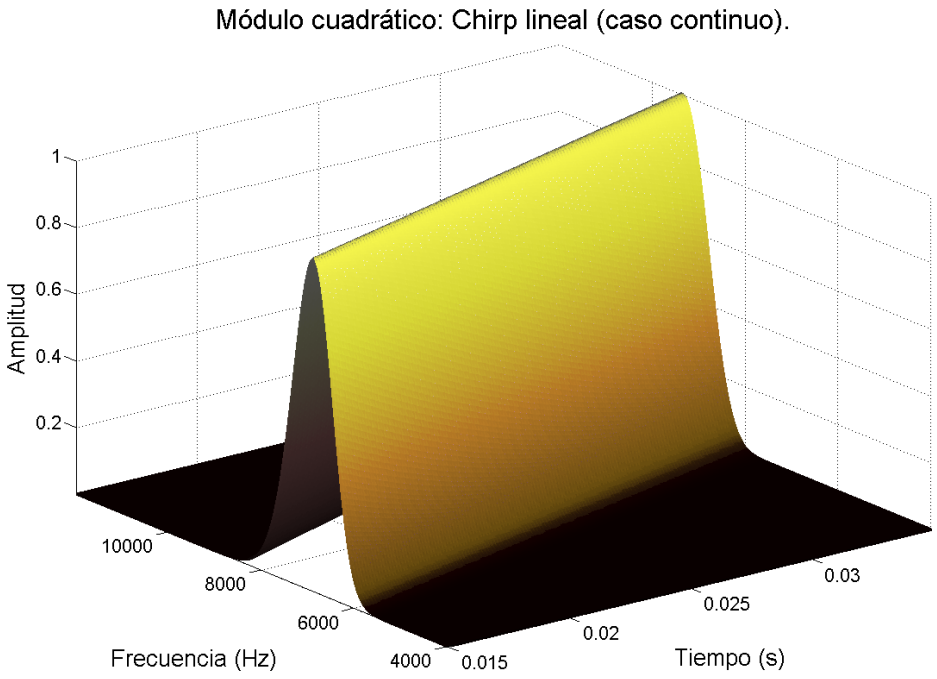


Figura 2.7: *Módulo cuadrático de los coeficientes wavelet para un chirp lineal de parámetros α =62500, β =5000 y γ =20, y amplitud constante $A(t) = A_1 = 1$.*

2.4. Metodología (II): paso al discreto

Por lo visto hasta ahora, los coeficientes wavelet llevan la información de amplitud y fase instantáneas (y por lo tanto, frecuencia instantánea) de la señal analizada incluso en las condiciones más generales, salvo ciertas correcciones, o funciones de penalización (en principio de segundo orden). Estas deducciones extraídas del desarrollo matemático detallado aquí no añaden nada nuevo a las de la literatura, si bien se ha obtenido una función matemática que proporciona la forma explícita de los coeficientes wavelet para unas condiciones locales asumibles en la mayoría de los casos. En las matemáticas desarrolladas en este capítulo, se encuentran codificadas las claves para superar los problemas del algoritmo [17], mencionados en varias ocasiones, que parecen no encajar con ninguna de las conclusiones destacadas.

Un diagrama de bloques resumido del algoritmo desarrollado por el Grupo de Audio Digital de la Universidad de Zaragoza aparece en la Figura 2.8.

El algoritmo calcula de forma automática, para cada señal de entrada, qué valores de frecuencia máximo y mínimo son necesarios para el análisis ($f_s/2$ y el mínimo entre la frecuencia ligada a la ventana o a la señal, y 20Hz, respectivamente), cubriendo a continuación de forma adecuada el eje de frecuencias, con un banco de filtros caracterizado por un número de divisiones por octava controlable por el usuario (por lo demás, de estructura similar al presentado en la Sección 3.3.3, excepto en el hecho de que en este caso el número de divisiones por octava es un escalar). A continuación se procede a llevar a cabo el análisis de la señal en sí, en dos barridos: en el inicial (grueso) quedan marcados los parciales sinusoidales que componen la parte más armónica de la señal. En el segundo (fino) se procede a la localización y análisis de los transitorios. En cada barrido se localizan las escalas para las que el escalograma del módulo de los coeficientes wavelet presenta máximos locales (crestas). Recogiendo la información de módulo y fase de la matriz compleja en tales bandas (esqueletos), se reconstruye perfectamente la frecuencia instantánea de los parciales, pero no así su módulo. Como se ha adelantado, se requiere una renormalización adicional, al valor de la señal. Es esta renormalización la que no parece encajar con lo predicho por la literatura, de la cual parece desprenderse que el proceso de recuperación de la amplitud instantánea utilizando crestas y esqueletos es independiente de $x(t)$.

El algoritmo presentado en la Figura 2.8 se enfrenta a situación peor en el caso de que la ley de variación frecuencial de la señal de entrada sea oscilante. Incluso con variaciones frecuenciales relativamente pequeñas, una señal monocromática de amplitud constante resulta reconstruida con un evidente rizado. Este problema parece más preocupante puesto que ninguna renormalización, salvo una variable con el tiempo, es capaz de eliminar un rizado. De no resolverse adecuadamente esta importante contingencia, cabe la posibilidad de que nos encontremos ante el límite de la técnica.

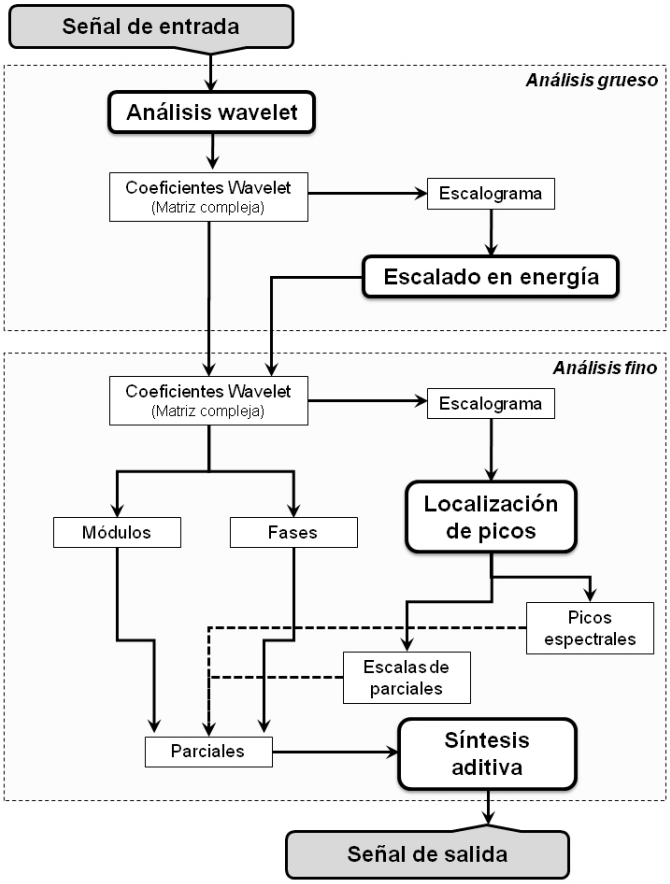


Figura 2.8: *Diagrama de bloques del algoritmo inicial.*

2.4.1. Datos experimentales iniciales e interpretación

La teoría posibilita en principio recuperaciones muy precisas en amplitudes y frecuencias para cada parcial de la señal. La práctica entraba en aparente conflicto con tales resultados. Así pues, había algún detalle incongruente entre la teoría y la práctica.

A la par que se iba profundizando en la obtención matemática de los coeficientes wavelet para señales cada vez más complejas, se fueron realizando pruebas con la versión inicial del algoritmo, aplicándolo sobre una serie de señales sintéticas de características conocidas. Analizando una señal monocromática de amplitud $A = 1$ y frecuencia determinada, en algunas ocasiones el algoritmo conseguía extraer el dato exacto de la amplitud, mientras que en otras la amplitud obtenida era inferior a la esperada. Parte de los resultados obtenidos se muestran en la Tabla 2.1 (para más información, véase el Anexo I.c):

Frecuencia	Amplitud original	Amplitud detectada	Ratio
40	1.0000	0.9873	1.0129
80	1.0000	0.9989	1.0011
160	1.0000	1.0000	1.0000
320	1.0000	1.0000	1.0000
640	1.0000	0.9299	1.0753
1280	1.0000	0.9299	1.0753
2560	1.0000	1.0000	1.0000
5120	1.0000	0.9767	1.0238
10240	1.0000	0.9328	1.0721
20480	1.0000	0.9553	1.0468

Tabla 2.1: *Resultados empíricos iniciales del análisis en amplitud de una señal de amplitud constante.*

2.4.1.1. Solución a la normalización

La interpretación a priori de los datos presentados en la Tabla 2.1 es complicada. Sin embargo, uniendo a tales datos la posición de cada uno de los filtros del banco para cada una de las señales analizadas (un resumen de cuyos datos presentamos en las figuras del Anexo I.c) se puede arrojar algo de luz sobre lo que está sucediendo.

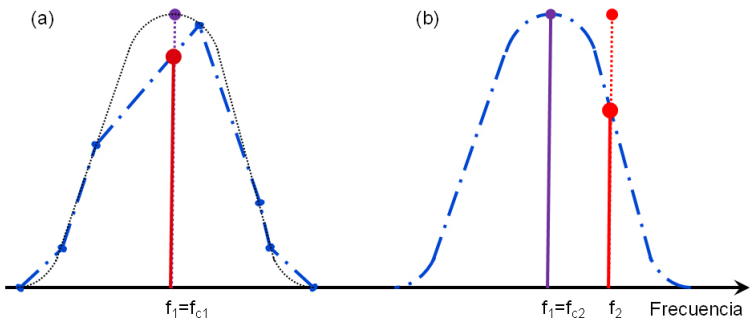


Figura 2.9: *Representación gráfica de los dos tipos de errores encontrados. (a) Falta de puntos de referencia. (b) Posición de los filtros.*

En la Figura 2.9 se ha representado de forma esquemática el resultado del análisis de una señal de frecuencia f_1 (f_2) bajo un cierto banco de filtros. En cada una de las subfiguras aparece la frecuencia buscada y el filtro más cercano a ella.

Se están produciendo dos errores de naturaleza muy diferente cuyo resultado aparente es el mismo. En primer lugar, tengamos en cuenta que para las frecuencias más bajas, cada banco de filtros supuestamente gaussiano está calculado en un limitado conjunto de puntos, como se ha intentado representar gráficamente en la Figura 2.9 (a). En este caso, al intentar localizar una frecuencia de valor f_1, no importa si existe o no un filtro del banco cuya frecuencia central esté situada justo sobre f_1, ya que la poca cantidad de información disponible arrojará un resultado sobre la amplitud de la componente (marcada con un punto grueso de color rojo) que no coincidirá con la esperada. Este problema se hace cada vez menos relevante a medida que se trabaja con frecuencias mayores, puesto que pronto se dispone de un gran número de puntos de cálculo para cada filtro. En tal caso, como se ha representado en la Figura 2.9 (b), el error proviene de si existe o no un filtro cuya frecuencia central coincida con la frecuencia buscada. Como se puede apreciar, cuando se intenta localizar una frecuencia $f_1 = f_c$, el resultado es la amplitud correcta. Sin embargo, cuando se trata de encontrar la amplitud asociada a la frecuencia f_2, sobre la que no hay centrado ningún filtro, se obtiene una amplitud menor de la esperada (marcada de nuevo en rojo).

Volviendo a los datos de la Tabla 2.1, los errores para las frecuencias de 40 y 80Hz son debidos al caso mostrado en la Figura 2.9 (a), mientras que el resto de errores encajan en lo mostrado en la Figura 2.9 (b). En el caso de las frecuencias de 160, 320 y 2560Hz, se dispone de un filtro calculado en un conjunto de puntos suficiente, centrado en cada una de ellas, de ahí que la amplitud rescatada sea exacta. En el Anexo I.c se han representado gráficamente los datos que revelan las conclusiones aquí presentadas.

Para estos problemas, la solución parece evidente. Por un lado, podemos calcular la gaussiana que pasa por un cierto número de puntos (con 3 es suficiente), lo cual permite reconstruir los filtros de los que se conocen pocos puntos con tanta resolución como sea necesario, y en concreto saber dónde se encontraría situada su frecuencia central, es decir, dónde se encuentra el máximo *teórico* del filtro, que no tiene por qué coincidir con el máximo práctico como se puede apreciar en la Figura 2.9 (a), línea negra punteada. A partir de aquí, se puede proceder de forma idéntica en ambos casos para reconstruir la señal.

En la Figura 2.10 aparece un ejemplo explicativo de la respuesta del banco de filtros a la localización de una frecuencia determinada (llamada aquí de nuevo f_1). Los filtros que ofrecen respuesta a tal frecuencia son los situados en su proximidad, en la Figura 2.10 (a), Ψ_2 a Ψ_5. La respuesta del resto puede considerarse nula a efectos prácticos, aunque tratándose de gaussianas no lo sea matemáticamente. Lo que debería idealmente ser una delta de Kronecker se convierte en teoría en una nueva gaussiana (según se ha explicado en la Sección 2.3), de la que debido a la discretización del banco de filtros, sólo se conocen unos pocos puntos. Esto provoca un error evidente en la obtención de la amplitud, como ya se ha dicho, además de una (experimentalmente pequeña) desviación en la localización de la

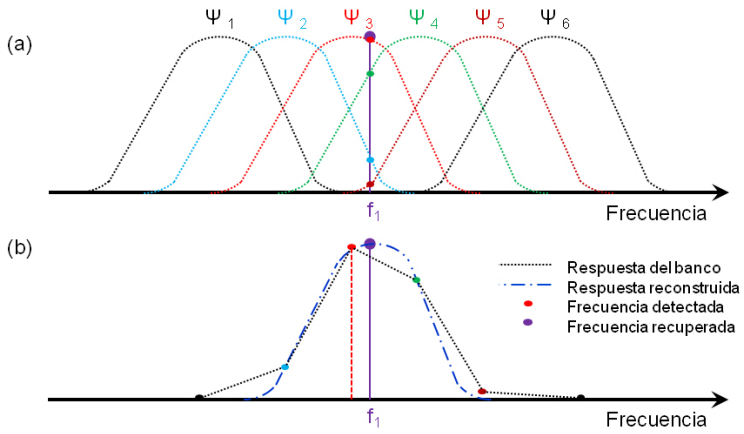


Figura 2.10: *Proceso de recuperación de la información. (a) Respuesta del banco de filtros. (b) Reconstrucción de la información.*

frecuencia instantánea, ambos datos calculados a partir del punto de máxima respuesta del banco de filtros, en rojo en la Figura 2.10 (b). Sin embargo, utilizando tres puntos de esta respuesta (línea negra punteada) se puede calcular cuál es la gaussiana que mejor encaja con los datos experimentales (curva azul discontinua). Con la expresión de la gaussiana así obtenida resulta factible resituar la frecuencia instantánea para cada parcial de la señal, así como afinar su amplitud (en morado en la figura), haciendo de este modo innecesaria la renormalización.

La discretización del banco de filtros es la responsable de la necesidad de renormalizar la señal. El procedimiento de afinamiento explicado resulta efectivo para evitar este problema, pero no así el rizado. Bajo esta nueva luz, y asumiendo que en la práctica resulta inviable conseguir la variación constante y continua de las variables tiempo b y escala a, el desarrollo teórico clásico resulta definitivamente inapropiado, y sus conclusiones no son aplicables al caso discreto.

El muestreo de la variable temporal b, como es habitual, no representa un problema excesivo. Sin embargo, ¿qué sucede cuando se analiza una señal de frecuencia variable, como por ejemplo el chirp lineal, bajo un banco de filtros discreto en frecuencia?

2.4.2. El proceso de discretización

Por lo visto hasta ahora, parece evidente que la posición de los filtros en el análisis algorítmico es pieza clave en la no obtención de la amplitud instantánea de la señal. La siguiente cuestión es resolver si sucederá algo parecido con el rizado.

Tras obtener la expresión matemática de $W_x(a, b)$ para el caso del chirp lineal, Ecuación

(2.64), se tuvo por vez primera acceso a la posibilidad de visualizar qué sucedía con los coeficientes al discretizar el banco de filtros en frecuencia. En la Figura 2.11 aparece representado el módulo cuadrático del mismo chirp lineal representado en la Figura 2.7, esta vez analizado con un banco de filtros *discreto* (en concreto de 16 divisiones por octava). El máximo de los coeficientes wavelet se encuentra localizado en las mismas frecuencias que en el caso continuo, pero si se sigue su trayectoria en el semiplano, se obtiene un rizado muy pronunciado.

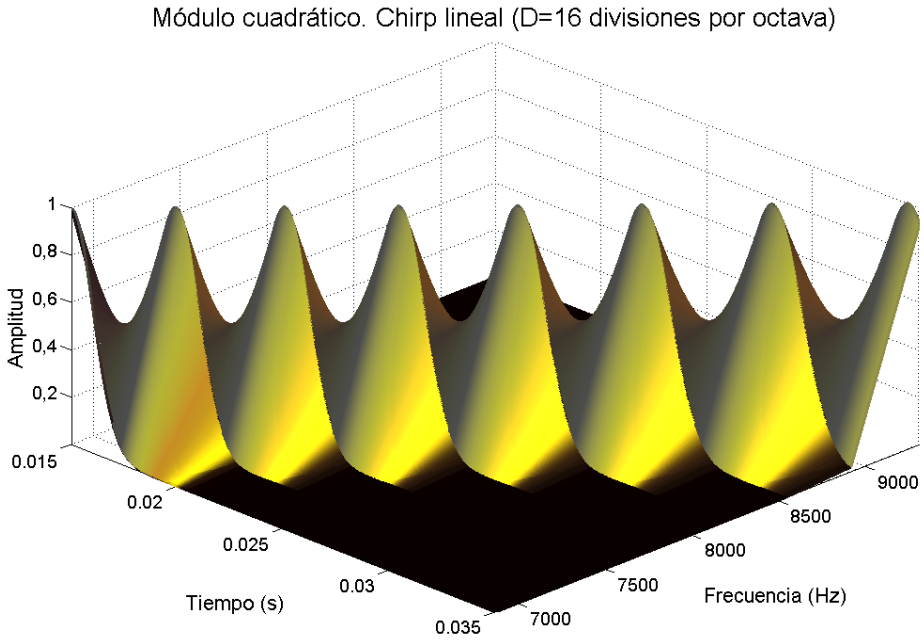


Figura 2.11: *Módulo cuadrático de los coeficientes wavelet para el mismo chirp lineal de la figura anterior, de parámetros α =62500, β =5000 y γ =20 y amplitud constante $A(t) = A_1 = 1$. El análisis bajo un banco de filtros real ($D = 16$ divisiones por octava, en este caso, o σ =0.3059) provoca un rizado muy evidente en el resultado final del análisis.*

A la vista de esta representación gráfica del proceso de discretización de la variable frecuencial, el origen del rizado queda perfectamente claro. En este punto, caben dos posibles conclusiones:

1. Se trata de un límite práctico de la técnica, con lo que se debería:
   - Cambiar la wavelet de análisis por otra con menor redundancia.
   - Abandonar la idea de la CCWT, en cuyo caso habría que buscar alternativas en principio más lógicas, como la Transformada Wavelet Discreta (DWT).

2. Existe algún modo de superar esta limitación.

En paralelo a la búsqueda de una respuesta definitiva a esta cuestión, se llevó a cabo un análisis detallado de posibles soluciones alternativas. Entre otras, se han valorado (y finalmente descartado, ya sea por su difícil implementación, la imposibilidad de resolver los coeficientes wavelet analíticamente o su prohibitivo tiempo de procesado intrínseco) la Transformada Wavelet Discreta [88], la Transformada Wavelet por Árbol Dual (Dual-Tree Wavelet Transform) [147] y la Transformada Chirplet [114].

Asimismo, se ha evaluado la posibilidad de utilizar una wavelet madre ortonormada, ya que de existir, permitiría una reconstrucción aditiva mucho más simple e intuitiva. En este aspecto, se han estudiado los Filtros de Espejo en Cuadratura de Respuesta Finita al Impulso (*Finite Impulse Response Quadrature Mirror Filters*, o FIR QMF) [3], y las bases wavelet generadas mediante pares de Hilbert de funciones [146, 158]. Quizá sea en este punto donde mayores probabilidades de éxito se han encontrado. Partiendo de la información básica acerca de bases wavelet ortonormadas [50], se ha valorado y trabajado fuertemente, de hecho, en la posibilidad de construir un banco de filtros ortogonal a partir de la propia wavelet de Morlet. La idea es que, anidando una gaussiana dentro de otra, el resultado es una función con un decaimiento más vertical. Un anidado de cuarto o quinto orden puede conseguir una cuasi-ortogonalidad práctica en la wavelet de Morlet, si bien emplear este nuevo banco de filtros tampoco elimina el rizado. Siguiendo en esta línea, se ha estudiado con cierta profundidad el trabajo de Vetterli y Kovačević acerca de bancos de filtros de reconstrucción perfecta [96, 97, 163], a veces relacionados con la propia Transformada Wavelet [98, 99]. Sin embargo, dado que finalmente se ha regresado a la wavelet de Morlet original superando sus limitaciones, este sin duda interesante tema se convierte en una rama de importancia bastante limitada.

La revisión bibliográfica en torno a estos puntos ha sido amplia y, si bien en su momento se valoraron cada una de las diferentes alternativas, todas ellas presentaban problemas adicionales.

2.4.2.1. Solución al rizado

De todo lo anterior se deduce que la superación final de este problema no ha resultado ni tan rápida ni tan evidente como la de la renormalización. La solución definitiva surge a partir de comprobar que la integral de los coeficientes wavelet en el eje de escalas da como resultado la obtención del par canónico de la señal, Ecuación (2.66), con un factor de sobrepeso determinado.

Como ya se ha dicho, integrar en un espacio continuo es equivalente a sumar en un espacio discreto. Por lo tanto, sumar los coeficientes wavelet en el espacio de las escalas discretas podría dar como resultado una aproximación al par canónico de la señal. El factor

de sobrepeso añadido, llegado el caso, es un precio más que aceptable, siempre que se consiga eliminar el rizado de salida.

En efecto, se ha comprobado que la suma de la información en las bandas asociadas a un parcial (por ejemplo los coeficientes wavelet complejos de los filtros Ψ_2 a Ψ_5 de la Figura 2.10), da como resultado una función muy próxima al par canónico de ese parcial (ver Anexo II.d.1.2). Esto permite una normalización mucho más simple que el método propuesto en la Sección 2.4.1.1 y por ende prácticamente elimina el rizado en la reconstrucción. Los detalles al respecto serán explicados en el siguiente capítulo.

2.5. Conclusiones y contribuciones

En este Capítulo se ha llevado a cabo un análisis matemático novedoso y riguroso de los coeficientes wavelet bajo ciertas aproximaciones (asumibles en la mayoría de los casos). La pérdida de generalidad respecto a la literatura original [41, 42, 51, 72, 76, 100, 101] se compensa al obtenerse funciones de variable compleja cuyo módulo y fase pueden ser estudiados de forma exacta, e incluso representados gráficamente. Con este análisis se pretende obtener el origen último de las aparentes incongruencias entre las previsiones teóricas y los resultados prácticos de la CCWT, con la idea de proponer soluciones a los mismos y poder mejorar la técnica de cara a sus posibles aplicaciones.

Las contribuciones presentadas en este bloque pueden dividirse en dos grandes grupos. En el primero de ellos, acerca del desarrollo matemático expuesto:

1. Inserción de un parámetro de control en la wavelet de Morlet, de cara a flexibilizar el banco de filtros generado.

2. Obtención de los coeficientes wavelet para casos resolubles analíticamente.

   - Señal monocromática de amplitud y frecuencia constantes.
   - Señal multicomponente de amplitudes y frecuencias constantes.
     - Análisis de los términos de intermodulación.
   - Señal monocromática de frecuencia constante y amplitud variable (envolvente gaussiana).
   - Señal monocromática de frecuencia variable (de aproximación cuadrática) y amplitud variable (cuasiestática).

3. Análisis de la integral en escalas de los coeficientes wavelet.

   - Caso: señal monocromática de frecuencia cuadrática y amplitud cuasiestática.

En el segundo grupo, sobre la base del mencionado desarrollo matemático, se ha procedido a explicar las limitaciones del algoritmo inicial, propuesto en [17], encontrando además las soluciones a las mismas:

1. Obtención del origen de la renormalización en el algoritmo inicial (espaciado y composición puntual del banco de filtros).
    - Solución: métodos alternativos de renormalización genérica.
2. Obtención del origen del rizado en el algoritmo inicial (cuantización del banco de filtros).
    - Solución: sumatorio de los coeficientes wavelet complejos en el eje de escalas.

Con estas contribuciones, se superan las limitaciones del algoritmo en su estadio inicial, demostrándose que la CCWT puede ser una herramienta adecuada para el análisis y la caracterización de las señales de audio. A partir de este momento, se llevará a cabo la optimización del banco de filtros y la puesta a prueba de la técnica en varias aplicaciones diferentes.

Capítulo 3

El Algoritmo C.W.A.S.

Índice

"Son vanas y están plagadas de errores
las ciencias que no han nacido
del experimento, madre de toda certidumbre".
Leonardo Da Vinci (1452-1519).
Pintor, escultor e inventor italiano.

A continuación, se va a exponer detalladamente el algoritmo para el análisis de las señales de audio que se ha derivado a partir de los resultados del capítulo anterior: el algoritmo de Síntesis Aditiva por Wavelets Complejas, o C.W.A.S. (en adelante CWAS), por sus siglas en inglés (Complex Wavelet Additive Synthesis). Se revisarán las diferentes aportaciones en torno a la Wavelet de Morlet, al diseño del banco de filtros generado mediante esta familia de funciones, al modelo subyacente de la señal de audio que arroja este análisis pasobanda, al proceso de renormalización de la información obtenida, así como a las diferentes técnicas de corte en bandas y seguimiento de parciales que se han ensayado y sus respectivos resultados.

3.1. Introducción

Como se concluyó al final del Capítulo 2, el desarrollo matemático presentado fue la base científica y la inspiración por la cual se encontraron las limitaciones del algoritmo original, pudiendo finalmente superarse. En este Capítulo, se van a detallar los cambios introducidos en la técnica de análisis. Puede dividirse en dos grandes bloques. El primero versará sobre las mejoras introducidas en la propia obtención de los coeficientes wavelet (parte de ello ya ha sido estudiado en la Sección 2.2, donde se detallaban los cambios introducidos en la wavelet de Morlet). Partiendo de estos cambios, se tratará sobre la generación del propio banco de filtros de análisis para incluir el proceso de renormalización automática.

El segundo bloque conceptual del Capítulo está orientado al tratamiento de la información contenida en la matriz de coeficientes wavelet, incluyendo la selección de las bandas frecuenciales asociadas a cada parcial y el seguimiento de los mismos a lo largo del eje temporal. La información así procesada desemboca en un nuevo modelo de la señal de audio, muy intuitivo para ciertas aplicaciones, y un tanto más abstracto de cara a otras, como se demostrará en capítulos posteriores.

3.2. Diagrama de bloques del algoritmo CWAS

A partir de las ideas expuestas al final del anterior capítulo, la técnica de análisis original ha ido evolucionando hasta el algoritmo CWAS presentado en este trabajo. Se han incluido cambios de mayor o menor calado en prácticamente todos los estadios del análisis de la señal de audio, desde el propio filtrado pasobanda a las aplicaciones posteriores, pasando por el procesado de la información. Un bloque de alto nivel del algoritmo queda reflejado en la Figura 3.1.

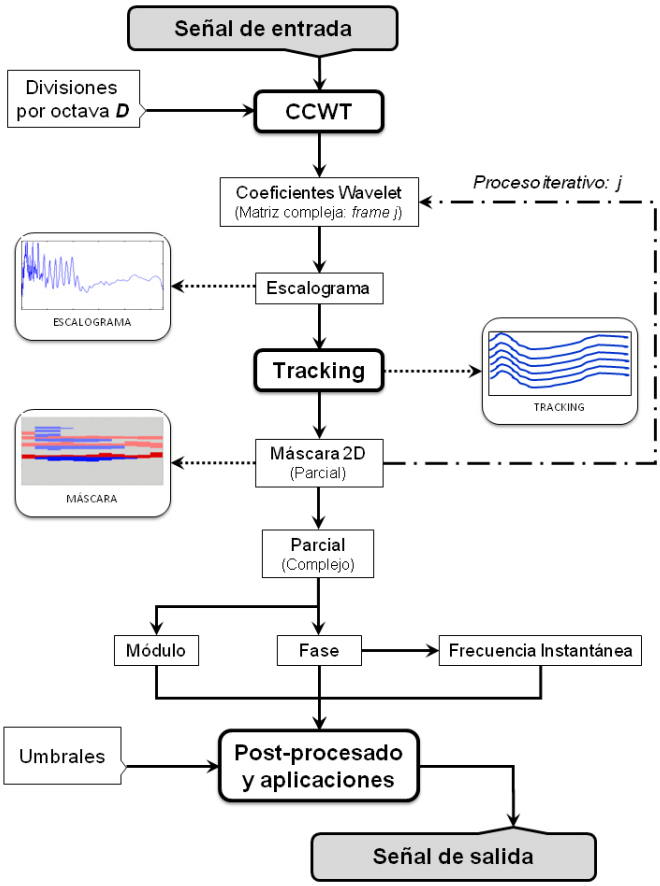


Figura 3.1: *Bloque esquemático de alto nivel del algoritmo CWAS.*

De forma muy resumida, el usuario tiene control sobre la resolución del sistema al poder escoger el número de divisiones por octava *para cada octava del espectro*. Con este simple control externo se calculan las contribuciones globales de los filtros y se renormalizan los datos a medida que estos se obtienen, de forma dinámica; así, pese a que ni la wavelet de

Morlet ni por lo ende la propia CCWT tal cual se han utilizado presentan carácter unitario, los coeficientes wavelet calculados sí conservan la energía de la señal de entrada. El módulo de los coeficientes está íntimamente ligado al escalograma de la señal (el cual se puede obtener de forma global, cada cierto número de muestras e incluso punto por punto), mediante el que es posible llevar a cabo un seguimiento de parciales. Tras este estadio, la información de salida tiene la forma de un conjunto de máscaras bidimensionales que, aplicadas sobre los coeficientes complejos originales, proporcionan como resultado la amplitud y fase instantáneas para cada uno de los parciales aislados detectados por el banco de filtros. Lo que se haga con esta información (obtenida de forma automática y completamente independiente de la señal de entrada) dependerá de la aplicación concreta que se quiera ejecutar a continuación. Algunas de las posibilidades a este respecto se detallarán en los Capítulos 4 y 5.

3.3. Banco de filtros

En esta Sección se va a detallar el desarrollo del banco de filtros de análisis a partir de la wavelet de Morlet revisada. Una de las cuestiones más importantes e interesantes a la hora de generar un banco de filtros pasobanda aceptable, es la correcta elección de la resolución del análisis y la subsiguiente distribución de la familia de filtros a lo largo del eje frecuencial. En este caso, se pretende construir un banco de filtros de propósito general, pero a la vez lo suficientemente flexible como para poder ser adaptado sin excesivas complicaciones a un abanico de aplicaciones tan amplio como sea necesario.

En general, se pretende que el análisis de la señal conserve las propiedades frecuenciales, energéticas, tímbricas y de duración de la señal original, es decir, que la señal de entrada y la señal sintetizada a partir de los coeficientes wavelet resulten tan parecidas como sea posible (idealmente, que no puedan ser distinguidas acústicamente). Por otro lado, las diferencias numéricas entre ambas deben resultar asimismo despreciables. La wavelet madre escogida presenta un comportamiento logarítmico en frecuencia (recordemos que el banco de filtros que genera es de Q constante) que resulta especialmente adaptable al comportamiento del oído humano. Esto permite obtener resultados sonoros de cierta calidad incluso utilizando resoluciones frecuenciales relativamente bajas.

En este trabajo se ha buscado un compromiso entre la calidad final de la resíntesis y la adaptabilidad del algoritmo, o, en otras palabras, entre la precisión en la separación frecuencial y el tiempo de computación (uno de los "caballos de batalla" más importantes que presenta la técnica propuesta).

3.3.1. Características iniciales

El banco de filtros que se va a utilizar como familia pasobanda de análisis es la wavelet de Morlet, expresada en una base diádica (y por extensión, discreta) k, en lugar del parámetro a continuo utilizado en el capítulo anterior. Esto supone que, en el dominio de la frecuencia, se puede escribir:

$$\hat{\psi}_k = Ce^{-\frac{\sigma^2\omega_0^2}{2}(\frac{k}{k_n}-1)^2} \tag{3.1}$$

con:

$$k_n = k_{min}2^{\frac{n}{D}}, \qquad n = 1, \ldots, J \cdot D \tag{3.2}$$

Así, cuando $D = 1$, el espectro queda dividido en J octavas. La escala mínima k_{min} está relacionada con la frecuencia máxima del análisis, f_{max}, y f_{max} a su vez con la frecuencia de muestreo f_s a través del criterio de Nyquist, $f_{max} = f_s/2$.

El conjunto total K de las escalas discretas puede escribirse como:

$$K = \left\{k_n\right\} \qquad n = 1, \ldots, J \cdot D \tag{3.3}$$

Parece evidente que el número de divisiones por octava D y el ancho de banda de los filtros, σ, deben estar de alguna forma relacionados si se pretende cubrir adecuadamente el eje frecuencial con la estructura del banco de filtros. Para resaltar la importancia de esto, en la Figura 3.2 se ha representado gráficamente una estructura de filtros pasobanda que no cumple con las relaciones ideales entre D, Q y σ que se obtendrán en la siguiente Sección, y que por lo tanto presenta algunos problemas en la adecuada cobertura del espectro. Esto es debido a que los filtros están demasiado separados entre sí (son demasiado estrechos). En una situación un poco más extrema, podría darse en caso de que algunas zonas frecuenciales del espectro no estuviesen cubiertas por *ningún* filtro del banco. En la Figura, los valores representados son $D = 1.5$, $Q = 2.2024$, que junto con el valor de ancho de banda $\sigma = 0.4$ proporcionan una contribución global (línea negra punteada) con evidentes altibajos. En las zonas de máxima atenuación, aunque no tiene por que ser el caso, podrían presentarse ciertos problemas.

A continuación, se van a obtener las expresiones que relacionan a D con Q y σ. De esta forma se pretende que, definiendo únicamente la precisión del filtrado a través de D, toda la estructura pasobanda del banco de filtros quede unívocamente determinada.

3.3.2. Factor de calidad, ancho de banda y divisiones por octava

Es primordial calcular en primer lugar los puntos de corte entre dos miembros consecutivos del banco de filtros. Como se ha avanzado, se trabajará con filtros pasobanda discretos,

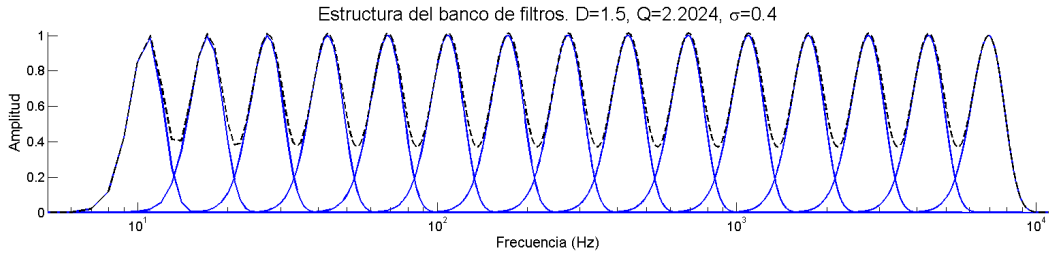


Figura 3.2: *Estructura de un banco de filtros de cobertura inadecuada. Linea punteada: contribución global.*

controlados por un factor de escala k_j de la forma (tomando $C = 1$):

$$\hat{\psi}_{k,j}(\omega) = e^{-\sigma^2 \frac{(k_j\omega-\omega_0)^2}{2}} \tag{3.4}$$

El ancho de banda del filtro $j - simo$ (marcado por las frecuencias que presentan un decaimiento de -3dB respecto del valor máximo del filtro) se puede obtener fácilmente de la Ecuación (3.4), siendo:

$$BW_j = \frac{2\sqrt{\ln 2}}{\sigma k_j} \tag{3.5}$$

Los puntos de corte entre el filtro $j - simo$ y el $(j+1) - simo$ se localizan en los valores del eje frecuencial ω donde las alturas de ambos son iguales, es decir, donde se cumpla:

$$e^{-\sigma^2 \frac{(k_j\omega-\omega_0)^2}{2}} = e^{-\sigma^2 \frac{(k_{j+1}\omega-\omega_0)^2}{2}} \tag{3.6}$$

Esto es, cuando:

$$(k_j\omega - \omega_0)^2 = (k_{j+1}\omega - \omega_0)^2 \tag{3.7}$$

La solución no trivial de la Ecuación (3.7) se da en los puntos:

$$\omega = \frac{2\omega_0}{k_j + k_{j+1}} \tag{3.8}$$

Uno de los grados de libertad de que se dispone, es la altura de corte para dos de tales filtros consecutivos del banco. En lo que resta, salvo indicación expresa de lo contrario, se ha tomado un decaimiento de nuevo de $-3dB$ $(1/\sqrt{2})$ de amplitud. Sustituyendo esta cantidad en la Ecuación (3.4), es decir:

$$e^{-\sigma^2 \frac{(k_j\omega-\omega_0)^2}{2}} = \frac{1}{\sqrt{2}} \tag{3.9}$$

es posible, utilizando la Ecuación (3.8), despejar la anchura σ como:

$$\sigma = \frac{\sqrt{\ln 2}}{\omega_0} \frac{k_{j+1} + k_j}{k_{j+1} - k_j} \tag{3.10}$$

El factor de calidad Q propio del banco de filtros es la relación entre la frecuencia central de uno cualquiera de sus miembros y el ancho de banda del mismo, es decir:

$$Q = \frac{\omega_{c,j}}{BW_j} \tag{3.11}$$

Utilizando la Ecuación (3.5) y teniendo en cuenta que:

$$k_j = \frac{\omega_0}{\omega_{c,j}} \tag{3.12}$$

Se obtiene la siguiente expresión para Q:

$$Q = \frac{\sigma\omega_0}{2\sqrt{\ln 2}} \tag{3.13}$$

En este punto, introduciendo esta expresión en la Ecuación (3.10) se puede concebir una relación entre el factor de calidad Q y los factores de escala de dos filtros consecutivos:

$$Q = \frac{1}{2} \frac{k_{j+1} + k_j}{k_{j+1} - k_j} \tag{3.14}$$

Y, utilizando de nuevo la Ecuación (3.12), se origina la expresión buscada para Q en función de las frecuencias centrales de dos filtros consecutivos:

$$Q = \frac{1}{2} \frac{\omega_{c,j} + \omega_{c,j+1}}{\omega_{c,j} - \omega_{c,j+1}} \tag{3.15}$$

Si $\omega_{c,j}$ es la frecuencia central del filtro $j - simo$, está relacionada con la escala $j - sima$ a través de la Ecuación (3.2). Las Ecuaciones (3.14) y (3.15) pueden ser obtenidas para el caso particular de $j = J - 1$ (recordemos que J es el número de octavas en las que se divide el espectro). En tal caso, Q puede ser expresado como:

$$Q = \frac{1}{2} \frac{k_J + k_{J-1}}{k_J - k_{J-1}} \tag{3.16}$$

A través de la Ecuación (3.13), se obtiene la relación entre el factor de calidad Q y el ancho de los filtros, σ, siendo:

$$\sigma\omega_0 = 2\sqrt{\ln 2}Q \tag{3.17}$$

Una vez escogido el valor de D, el conjunto K de las escalas discretas, se obtiene a través

de la Ecuación (3.3). Utilizando las Ecuaciones (3.16) y (3.17) respectivamente, se pueden computar Q y σ. El número de divisiones por octava define de esta forma la estructura completa del banco de filtros.

En la Figura 3.3 se representa una banco de filtros que cumple con las relaciones adecuadas de los tres parámetros, contrariamente a lo que sucedía en la Figura 3.2. En este caso concreto, $D = 2$, $Q = 2.9142$ y $\sigma = 0.2426$. Como se puede observar, la contribución global del banco de filtros (linea negra punteada) es mucho más regular que en el caso anterior. En la figura, el evidente rizado que se puede observar es debido principalmente al bajo número de D escogido.

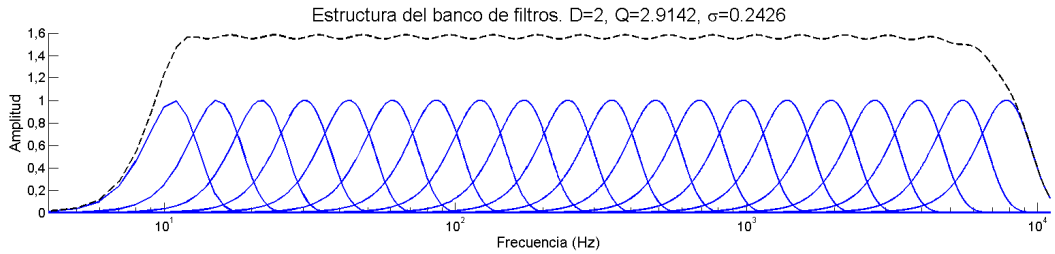


Figura 3.3: *Estructura de un banco de filtros de cobertura adecuada. Linea punteada: contribución global.*

Como se verá en la Sección 3.6, esta contribución global juega un papel determinante en el proceso de renormalización de la señal.

3.3.3. Estructura final del banco de filtros

Ahora que se dispone de una relación entre las frecuencias centrales de los filtros consecutivos cualesquiera y el factor de calidad Q que garantiza una altura de corte para los filtros de $-3dB$, es decisión del usuario elegir la adecuada resolución del filtrado, es decir, el valor de D. Menores valores del parámetro proporcionan menor resolución espectral y menor tiempo de cómputo.

En las secciones precedentes y hasta aquí, se ha supuesto que el número de divisiones por octava D que controla la resolución del filtrado es una cantidad constante. Sin embargo, esto puede resultar no solo innecesario sino incluso contraproducente. Supóngase, por ejemplo, un valor de $D = 16$ divisiones por octava. Esto supone colocar la misma cantidad (16 filtros) en cada una de las diez octavas del espectro. En otras palabras, se posicionan 16 filtros entre los 20 y los 40Hz, y un número idéntico para cubrir las frecuencias entre los 10kHz y los 20kHz.

Bajo este punto de vista parece evidente que, al menos a priori, es necesaria menor

precisión relativa en la zona de baja frecuencia, precisión creciente en la zona de frecuencias medias y una precisión más elevada en la zona de alta frecuencia, donde los filtros de Q constante resultan ser comparativamente demasiado anchos con respecto a sos homólogos de baja frecuencia.

Esto es exactamente lo que se lleva a cabo en la versión definitiva del algoritmo básico. Para cada una de las 10 octavas del espectro (suponiendo un análisis completo entre 20Hz y 20kHz aproximadamente) el usuario puede escoger un número de divisiones por octava (resolución) arbitrario. Dentro del análisis de una señal, el banco de filtros completo (y renormalizado según lo expuesto en la Sección 3.6.3), es calculado una vez y guardado en memoria, accediéndose a él siempre que sea necesario (lo cual tiende a optimizar el tiempo de cálculo al evitar repetir las mismas operaciones en cada iteración).

En el caso de trabajar con diferentes resoluciones en cada una de las octavas del espectro, el cálculo de los parámetros correspondientes al banco de filtros no puede ser global; estará necesariamente ligado a una octava concreta. La forma más razonable de proceder es trabajar en primer lugar con el filtro situado en el extremo superior del espectro, en la banda de menor escala (octava más alta, frecuencia máxima detectable). A través de la Ecuación (3.16) se puede conocer de forma inmediata el valor de su parámetro de calidad (el cual afectará a toda la octava por igual). Con este dato y a través de la Ecuación (3.17) es posible calcular el ancho de banda asociado (se tomará como parámetro de frecuencia de referencia $f_0 = 20Hz$). Este filtro límite está asociado a la escala mínima k_{min} de la Ecuación (3.2), lo que hace posible generar directamente cada una de las escalas (frecuencias) centrales de cada uno de los filtros asociados a la octava superior. El filtro de menor escala situado dentro de esta octava hace las veces de límite superior para la octava inmediatamente por debajo, de modo que, aplicando las Ecuaciones (3.16), (3.17) y (3.2) recursivamente, se puede generar el banco de filtros completo en todas las octavas del espectro de análisis.

Ya sea mediante un número de divisiones por octava constante o variable, una vez que la disposición en escalas y el ancho de banda de los filtros pasobanda quedan definidos, se pueden calcular los coeficientes wavelet de la señal procesada a través de las Ecuaciones (1.57) o (1.58). Los datos así obtenidos tendrán una estructura matricial determinada, la cual puede ser analizada en dos dimensiones (*espectrograma* wavelet) o en secciones unidimensionales (*escalograma* wavelet).

3.4. Matriz de coeficientes CWT: análisis en una y dos dimensiones

Al estar trabajando con una TFD de variable *compleja*, los coeficientes resultado son asimismo números complejos y por lo tanto se tiene acceso de forma simple y compacta tanto a la información modular como de fase para cada una de las bandas de análisis, como

se explicó en la Sección 1.5.2. En la mayoría de las ocasiones para lo que resta de disertación, se trabajará con un número cercano a las 200 bandas de frecuencia, bastante por debajo de la resolución que ofrece la STFT por ejemplo, pero con alguna ventaja adicional como la recuperación de la fase en cada punto del análisis utilizando para ello una cantidad de operaciones muy por debajo de las necesarias con la STFT para obtener el mismo resultado, como se demostrará en la Sección 5.2.

El tamaño exacto de la matriz de coeficientes complejos (*matriz CWT*) será BxM, donde B es el número de bandas de análisis y M el número de muestras temporales de duración de la señal. Por ejemplo, para una señal de clarinete de f_s = 22050Hz, t = 1.14 segundos de duración aproximada y una distribución de divisiones por octava de $D = [6; 8; 10; 12; 24; 32; 32; 32; 32]$[1] se obtiene una matriz coeficientes de tamaño 25228x189. Teniendo en cuenta que cada dato es un número complejo de 16 bits, si la señal original tiene un peso total de 49.3Kb, el tamaño final de la matriz de coeficientes (que, además, almacena una serie de datos de interés sobre la señal analizada como la ruta de acceso, frecuencia de muestreo, etc.) es de 36.3Mb. Con este volumen de datos, las aplicaciones en tiempo real de esta técnica deben ser abordadas mediante sistemas alternativos como la computación en paralelo o sistemas de aceleración por hardware (procesado a través de la GPU). Para los expertos, el algoritmo CWAS representa un atractivo desafío.

3.4.1. Evolución temporal de la información: Espectrograma wavelet

La matriz CWT puede ser estudiada en módulo y fase. Como se vio en el Capítulo 2, en el módulo de los coeficientes se encuentra codificada toda la información necesaria para caracterizar $x(t)$, si bien resulta prohibitivamente complicado extraer de ellos la información frecuencial (la cual será extraída a partir de la fase). Sin embargo, la ambivalencia de la información presente en el módulo lo convierte en una poderosa herramienta de visualización de resultados. A $\|W_x(k,t)\|$ de los coeficientes complejos se llamará en adelante el *espectrograma wavelet*. Se trata de una representación tridimensional de la información contenida en los coeficientes complejos. Esta información, con propósitos de claridad, será reducida en una dimensión y representada en el plano T–F. Por lo demás, la amplitud (eje Z) se simulará mediante un adecuado mapeo de color de los datos (concretamente, los tonos claros se corresponden con las amplitudes más elevadas). En la Figura 3.4 aparece representado el espectrograma wavelet de una señal musical (concretamente, un extracto de la pieza clásica *"Claros y frescos ríos"*[2]. En esta figura se pueden apreciar en color claro las trayectorias

[1] *Este es el valor de D utilizado por defecto, y salvo indicación expresa de lo contrario, para señales de f_s =22050Hz. $D = [4; 8; 12; 32; 32; 32; 24; 24; 16; 16]$ cuando f_s =44100Hz, para un total de 201 bandas frecuenciales. El programa de análisis selecciona automáticamente el parámetro D en función de la frecuencia de muestreo de la señal de entrada.*

[2] *Tres Libros de Música, Libro III*, de Alonso de Mudarra, Sevilla, 1546.

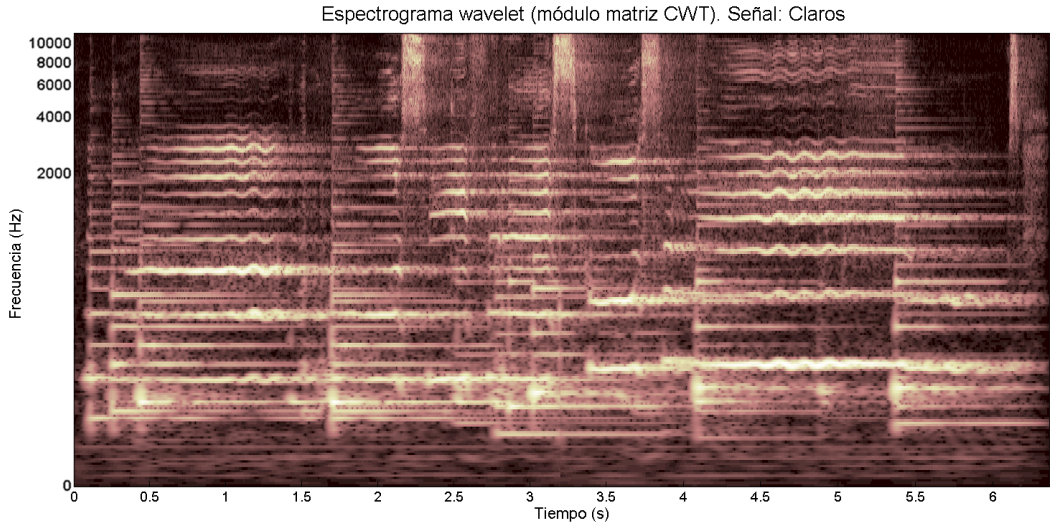


Figura 3.4: *Espectrograma wavelet de la señal "Claros". Las trayectorias de voz e instrumento aparecen claramente diferenciadas.*

frecuenciales (eje Y) correspondientes a las diferentes componentes sinusoidales detectadas y su evolución temporal (eje X). Las notas de la *vihuela*[3] aparecen como trayectorias rectilíneas de corta duración, mientras que la voz presenta un comportamiento mucho más armonioso y modulado. La distinción entre diferentes grupos de trayectorias frecuenciales será de capital importancia en la Separación Ciega de Fuentes de Audio (BASS), que se verá en el Capítulo 4.

3.4.2. Componentes espectrales: Escalograma

La intersección de la información representada en el espectrograma wavelet con los planos de $X = cte$ ($t = t_i \; \forall t_i \in T$, donde T es la duración temporal de la señal) representa el *escalograma* de la señal. Es una distribución de picos que representa la distribución energética del contenido frecuencial de $x(t)$ en el instante de tiempo t_i concreto del corte. La evolución de estos picos (las *crestas* de Morlet, Kronland-Martinet y Carmona) en el tiempo, es lo que define el *esqueleto* de la Transformada. Si se suman los datos correspondientes a varias de estas muestras, se obtiene el escalograma acumulado, que puede llegar a contener toda la información energética de la señal (sin más que sumar el módulo de los coeficientes en todos los instantes de tiempo de duración de la señal). En la Figura 3.5 aparecen tres escalogramas distintos, correspondientes a la misma señal presentada anteriormente. En la gráfica de la

[3] *La vihuela es un instrumento de cuerda que fue muy popular en la Península Ibérica (España) y en menor medida en Italia durante el siglo XVI. En España, coexistió con el laúd.*

izquierda, el escalograma instantáneo correspondiente a la muestra #1600. En el centro, el escalograma acumulado entre las muestras #1600 a #5000. Por último, a la derecha, el escalograma total de la señal. En las tres gráficas, la información vertical aparece en dB.

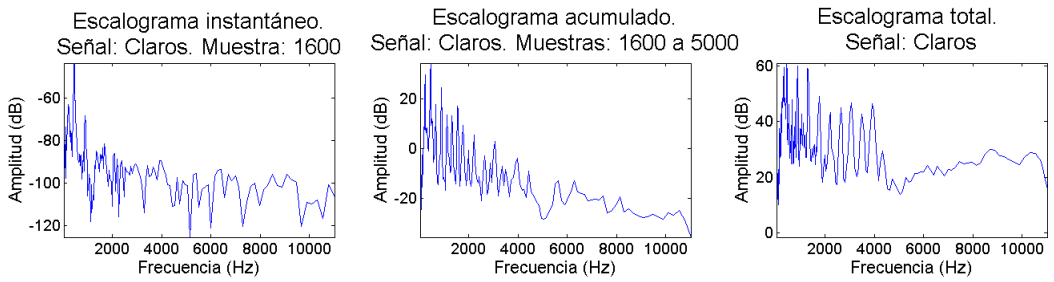


Figura 3.5: *Tres escalogramas relacionados con la señal "Claros". Instantáneo, acumulado y total.*

3.5. Modelo de la señal de audio

Implícito en la matriz CWT se encuentra un nuevo modelo para la señal de audio que presenta una gran coherencia tanto en el dominio del tiempo como en el de la frecuencia. Este modelo, basado en zonas de influencia y no en picos espectrales simples, resulta muy intuitivo para algunas aplicaciones y/o señales, mientras que en otras se puede calificar como simplemente *efectivo* (es decir, más abstracto). Este hecho quedará reflejado en los Capítulos 4 y 5.

Como ya se ha demostrado (Sección 2.4.2), la información contenida exclusivamente en los picos espectrales del escalograma (mucho más cercanos al concepto clásico de *parcial*) resulta insuficiente de cara a la coherencia final del modelo. Esto no es óbice para que la representación concatenada bidimensional de esta información (el espectrograma wavelet) resulte una de las herramientas visuales más poderosas de cara a comprender adecuadamente el modelo subyacente de señal.

3.5.1. Osciladores sinusoidales: nuevo concepto de Parcial

En efecto, si bien los parciales de $x(t)$ son localizados en principio a través de los máximos locales detectados en su escalograma (responsables de la distribución instantánea de picos), limitar el parcial a la información contenida exclusivamente en la(s) banda(s) donde se localiza tal máximo causaría, como se explicó en su momento, un evidente rizado en la reconstrucción de la señal. La solución a este rizado consiste en la suma de los coeficientes wavelet en una serie de bandas asociadas a cada máximo local.

En otras palabras, cada pico del escalograma tiene asociados un límite inferior $k_{inf,n}$ y otro superior $k_{sup,n}$ en la distribución de bandas de análisis, como se verá en detalle más adelante. El concepto de parcial adaptado al modelo propuesto consiste, paralelamente a lo propuesto en [107], en la acumulación de la información contenida en las coeficientes wavelet *complejos* dentro de estos límites, es decir:

$$p_n(b) = \sum_k W_x(k,b) \quad \forall k \epsilon K_n \tag{3.18}$$

En esta ecuación, b es la variable temporal muestreada (en adelante t), mientras que k representa, como se ha adelantado en la Sección 3.3.1, a la variable de escala discretizada. Por otro lado, K_n es el conjunto de escalas asociadas al parcial $n-simo$, p_n. Se cumple que:

$$K_n = \left\{ k \epsilon K : k_{inf,n} \geq k \geq k_{sup,n} \right\} \tag{3.19}$$

donde K es el conjunto total de escalas del análisis definido por la Ecuación (3.2), mientras $k_{inf,n}$ y $k_{sup,n}$ son respectivamente los límites inferior y superior relacionados con el pico $n-simo$ del escalograma (y por ende con p_n), sobre cuya obtención se tratará más en detalle en la Sección 3.8.

Calculado a través de la Ecuación (3.18), cada parcial detectado, $p_n(t)$ resulta evidentemente ser una función definida compleja, y por lo tanto referible en términos de módulo y fase a través de las Ecuaciones (1.43) y (1.44). Aplicando tales ecuaciones, es posible obtener la amplitud y la fase instantáneas para cada parcial, siendo:

$$A_n(t) = \|p_n(t)\| = \sqrt{\Re[p_n(t)]^2 + \Im[p_n(t)]^2} \tag{3.20}$$

y:

$$\phi_n(t) = arg\left(\Re[p_n(t)] + j\Im[p_n(t)]\right) \tag{3.21}$$

De este modo, cada *componente parcial* detectado de la señal puede escribirse como:

$$\rho_n(t) = A_n(t)\cos\left[\phi_n(t)\right] = \Re\left[p_n(t)\right] \tag{3.22}$$

Por otro lado, la frecuencia instantánea de $p_n(t)$ y por extensión de $\rho_n(t)$, se calcula a través de la Ecuación (1.45), obteniéndose:

$$f_{ins,n}(t) = \frac{1}{2\pi}\frac{d[\phi_n(t)]}{dt} \tag{3.23}$$

3.5.2. Síntesis Aditiva

Una vez procesados los N parciales detectados por el banco de filtros, que componen la señal analizada, un simple proceso de síntesis aditiva permite obtener una señal de salida $x_{syn}(t)$, siendo:

$$x_{syn}(t) = \sum_{n=1}^{N} \rho_n(t) = \sum_{n=1}^{N} A_n(t) \cos[\phi_n(t)] \tag{3.24}$$

Como se ha dicho, esta señal conserva las propiedades tímbricas, frecuenciales, energéticas y de duración de la señal original. La alta coherencia en ambos dominios del modelo subyacente permite obtener resultados de muy elevada calidad en la resíntesis. Más aún, la coherencia en fase obtenida permite calcular una señal de error $e(t)$, *restando las señales original y sintética en cada instante de tiempo*, es decir:

$$e(t) = x(t) - x_{syn}(t) \qquad \forall t \tag{3.25}$$

Como se verá en la Sección 5.5.2, y más en detalle en la Sección 5.5.4, la energía de esta señal de error es muy pequeña (de hecho puede perfectamente ser despreciada en la mayoría de los casos), y por lo tanto las señales original y sintética resultan prácticamente indistinguibles acústicamente.

3.6. Renormalización

Como se adelantó en la Sección 2.4.2.1, esta nueva definición de parcial, más alejada del concepto clásico, conlleva un detalle a tener cuidadosamente en cuenta: un cierto factor de *sobrepeso*. Este factor está relacionado con la respuesta del banco de filtros a la localización de cierta frecuencia (ver Figura 2.10). Dada la alta redundancia del filtrado pasobanda llevado a cabo, la detección de una componente determinada será no nula para múltiples gaussianas situadas en su vecindad frecuencial. Por lo tanto, tendremos información acerca de la componente de interés en un cierto número de bandas (contiguas). Como se explicó en la Sección 2.4.1.1, la respuesta del banco de filtros continuo es asimismo una gaussiana, mientras que en el caso discreto será una distribución con una cierta abertura, que presentará un pico a una frecuencia determinada. Al sumar los coeficientes wavelet encerrados bajo de una de estas distribuciones, se añade cada una de las contribuciones de cada una de las bandas involucradas, lo cual causa el citado sobrepeso. Como se verá inmediatamente, este parámetro es perfectamente conocido una vez definido el banco de filtros. De hecho, existen varias formas alternativas de tenerlo en cuenta.

3.6.1. Sobrepeso

En esta sección se pretende mostrar dos formas equivalentes de calcular el parámetro de sobrepeso que será utilizado como constante de normalización en el algoritmo CWAS. Para llevar a cabo tal cálculo, será suficiente estudiar el caso de una señal $x(t)$ de amplitud constante y frecuencia pura, en la cual se empleará el subíndice α para indicar que se trata de una variable real (es decir, f_α no tiene por qué coincidir con la posición de ninguno de los filtros del banco de análisis). Tal señal de entrada será de la forma:

$$x(t) = A_\alpha \cos(2\pi f_\alpha t) \tag{3.26}$$

la cual pretende ser analizada bajo el banco de filtros diádico de las Ecuaciones 3.1 y 3.2. Esto quiere decir que el $n - simo$ filtro del banco es una curva discreta de tipo gaussiano, centrada en la escala $k_n = \omega_0/\omega_n$. Matemáticamente, una gaussiana se anula sólo en el infinito, y por lo tanto la base del banco de filtros es no-ortogonal y la frecuencia f_α de la señal será detectada por todos los miembros del banco de filtros (véase de nuevo la Figura 2.10). Obviamente, la amplitud de la detección de f_α por el $n - simo$ filtro de la base es el valor del filtro evaluado en la escala $k_\alpha = \omega_0/\omega_\alpha$.

De esta forma, la detección global G_d de la frecuencia f_α será la suma de las contribuciones de todos los filtros de la base:

$$G_d = CA_\alpha \sum_{n=0}^{J \cdot D} e^{-\frac{\sigma^2 \omega_0^2}{2}(\frac{k_\alpha}{k_n}-1)^2} \tag{3.27}$$

En la práctica, la contribución principal a este sumatorio viene del filtro cuya frecuencia central es la más próxima a f_α, y los filtros situados en las inmediaciones de este:

$$G_d \approx CA_\alpha \sum_{n=n_{inf}}^{n=n_{sup}} e^{-\frac{\sigma^2 \omega_0^2}{2}(\frac{k_\alpha}{k_n}-1)^2} \tag{3.28}$$

donde $k_{n_{inf}} < k_m < k_{n_{sup}}$ y el $m - simo$ filtro es aquel que ofrece el valor máximo en la detección de f_α.

Sea λ:

$$\lambda = \sum_{n=0}^{J \cdot D} e^{-\frac{\sigma^2 \omega_0^2}{2}(\frac{k_\alpha}{k_n}-1)^2} \approx \sum_{n=n_{inf}}^{n=n_{sup}} e^{-\frac{\sigma^2 \omega_0^2}{2}(\frac{k_\alpha}{k_n}-1)^2} \tag{3.29}$$

Obviamente, tomando $C = 1/\lambda$ en las Ecuaciones (3.27) o (3.28), la detección global se puede escribir como:

$$G_d = CA_\alpha \lambda = A_\alpha \tag{3.30}$$

Se puede obtener el mismo resultado partiendo de la ecuación del módulo cuadrático de

los coeficientes wavelet:

$$\|W_x(a,b)\|^2 = \frac{A(b)^2C^2}{2}\pi\sigma^2a^2e^{-\sigma^2[\phi'(b)a-\omega_0]^2} \tag{3.31}$$

Para una señal de AM monocromática en la cual $A(t) = A_\alpha$ y $\phi(t) = \omega_\alpha t$, tomando:

$$C = \sqrt{\frac{2}{\pi}}\frac{1}{a\sigma}C' \tag{3.32}$$

la ecuación del módulo de los coeficientes en el caso discreto queda:

$$\|W_x(k,b)\| = C'A_\alpha e^{-\frac{\sigma^2}{2}(k\omega_\alpha-\omega_0)^2} \tag{3.33}$$

La detección total de la única componente viene del sumatorio del módulo en el eje discreto de escalas. Por lo tanto:

$$G_d = C'A_\alpha\sum_{n=0}^{J\cdot D} e^{-\frac{\sigma^2\omega_0^2}{2}(\frac{k_\alpha}{k_n}-1)^2} \tag{3.34}$$

expresión que coincide plenamente con la de la Ecuación (3.27). Esta igualdad se hace evidente si consideramos el origen de los coeficientes wavelet y el hecho de que las Ecuaciones (1.57) y (1.58) se han utilizado en los primeros párrafos de esta Sección para obtener la expresión del parámetro λ.

En una señal multicomponente, se presenta una contribución sobrepesada como la dada por la Ecuación (3.27) para cada parcial de la señal.

3.6.2. Resultados experimentales para el parámetro de sobrepeso

Utilizando este procedimiento, se han encontrado dos resultados clave:

- Los sumatorios de las Ecuaciones (3.27) o (3.34) convergen rápidamente a un resultado que depende principalmente de la estructura del banco de filtros de partida y por lo tanto del número de divisiones por octava D, pero que resulta independiente de la señal de entrada.

- Sumando los coeficientes wavelet complejos en todas las bandas que definen un parcial, como se ha repetido anteriormente, el rizado del proceso de cuantización prácticamente desaparece, sin que ello suponga ninguna pérdida de precisión en la detección de la frecuencia instantánea [31, 32]. El modelo de síntesis aditiva utilizado en la resíntesis es por lo tanto coherente en amplitud y fase para cada parcial detectado, como se demostrará en el Capítulo 5.

Los resultados experimentales del parámetro de sobrepeso λ obtenido, en función del número de divisiones por octava D se presentan en la Tabla 3.1.

Divisiones por octava D	**Parámetro de normalización** λ
4	1.5184
6	1.5124
8	1.5104
12	1.5090
16	1.5085
20	1.5082
24	1.5081
32	1.5080

Tabla 3.1: *Resultados empíricos del parámetro λ en función de D.*

3.6.3. Renormalización efectiva

Una vez más, existen pequeñas diferencias entre las expresiones matemáticas obtenidas de forma teórica y las utilizadas en la práctica. En este caso, el parámetro de sobrepeso que aparece en la Tabla 3.1 se considera *constante*, dado el número de divisiones por octava característico del banco de filtros. Pero, experimentalmente, presenta cierto rizado alrededor de los valores señalados en la tabla, oscilaciones que son tanto más pequeñas cuanto mayor sea el valor de D (como se adivina en las Figuras 3.2 y 3.3). Tales variaciones provienen de la contribución global del banco de filtros en cada instante de tiempo. Cuanto mayor es el número de divisiones por octava, menor resulta el factor de rizado y más próximo al dato constante experimental el resultado promedio de λ. Por otro lado, los cálculos presentados en la Sección 3.6.1 se han podido efectuar gracias a considerar constante el número de divisiones por octava, si bien en la versión definitiva del banco de filtros (Sección 3.3.3), esto no tiene por qué ser cierto.

Sin embargo, este hecho puede ser tomado en consideración sin excesivas dificultades. En efecto, dado un banco de filtros determinado, la contribución global de todos los miembros de la familia diádica queda perfectamente determinada para cada banda del análisis y en cada instante de tiempo, sin más que sumar los propios filtros entre sí (de forma equivalente a como se hizo para representar las contribuciones globales de los bancos de filtros de las Figuras 3.2 y 3.3). Así, dividiendo la contribución de cada banda y muestra por el valor del sobrepeso o contribución global evaluado para ella, se consigue la renormalización efectiva. La mejor forma de proceder es renormalizar los filtros del propio banco por el valor de su

contribución global en cada instante temporal *antes* de aplicar el filtrado pasobanda a la señal de entrada, como se verá a continuación.

3.7. Obtención de los coeficientes wavelet

De todo lo explicado hasta ahora, el único punto que permanece sin desarrollar es el modo exacto en que se obtiene la matriz CWT. En versiones previas del algoritmo, en las cuales se efectuaba el análisis de la señal completa en un único paso, el modo de operación era muy sencillo: a partir de D (constante o variable; no resulta especialmente importante distinguir en este punto) se pueden obtener los parámetros básicos del banco de filtros, y con ellos los propios filtros (adecuadamente normalizados). A continuación se aplica directamente la Ecuación (1.58), obteniéndose de este modo los coeficientes wavelet de $x(t)$.

Sin embargo, al aplicar este procedimiento directamente sobre una señal leída trama a trama, aparecen discontinuidades muy evidentes (y perfectamente audibles) en los límites entre frames. Por lo tanto, el modo de cálculo de los coeficientes debe ser adecuadamente revisado para superar estos problemas de borde. La solución consiste en aplicar la técnica clásica de la *convolución circular* [126] a la señal adecuadamente cortada y enventanada, tal como se detalla en [178].

En los siguientes apartados se detallará por un lado la lectura y enventanado secuenciales de la señal, así como la obtención final de los coeficientes wavelet de cada frame (mediante un proceso de *overlap–add*), y a continuación la convolución circular en sí misma, así como la obtención de la matriz de coeficientes complejos.

3.7.1. Overlap–add

Desde un punto de vista de alto nivel, el proceso resulta engañosamente sencillo. Aparece resumido en la Figura 3.6. De cara a obtener los coeficientes wavelet de un frame de N muestras (como se verá en la sección 3.9.3, el valor por defecto para el tamaño del frame es de $N = 4095$ muestras), se leerán $2N$ muestras de la señal de audio (guardada en formato `.wav`) adecuadamente extendida con un padding inicial de N muestras y un padding final de tamaño necesario para trabajar con un último frame de tamaño completo ($2N$).

Los datos leídos son suavizados (utilizando para ello una ventana de Hanning de $2N$ muestras) previamente al cálculo de sus coeficientes wavelet. Del resultado obtenido, se guarda la segunda mitad (enventanada por el fade-out de Hanning). A continuación, se leen las siguientes $2N$ muestras de la señal, comenzando a partir de la $N - sima$, y repitiéndose el proceso.

Por ejemplo, en la Figura 3.6, (parte izquierda) se han representado las primeras 8190 muestras de una señal real (un clarinete), mientras que en la parte derecha se ha tenido

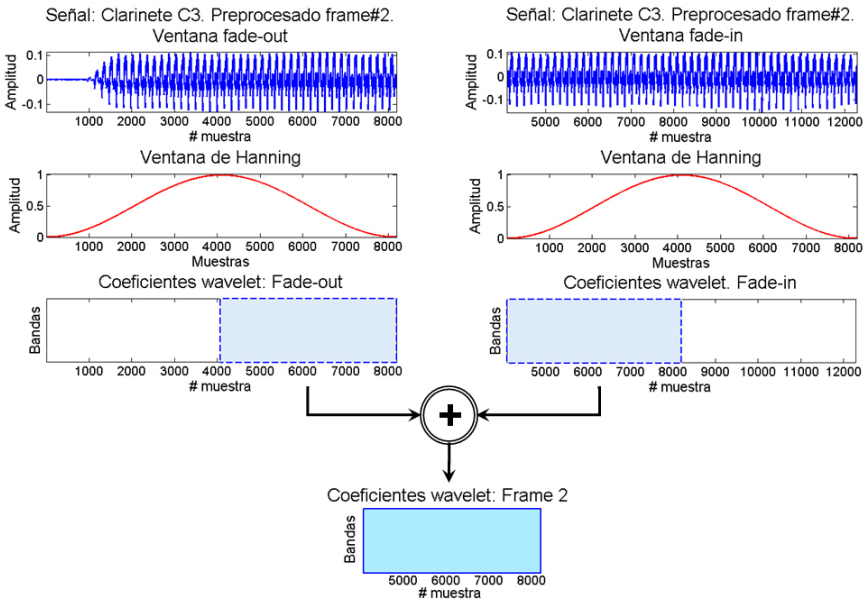


Figura 3.6: *Ejemplo de obtención de los coeficientes wavelet del frame $j = 2$ de una señal de clarinete, por medio de la técnica de overlap–adding.*

acceso a otras 8190 muestras de la misma, comenzando a partir de la 4096 (es decir, las muestras 4096 a 12286). Por lo tanto, la mitad final del primer proceso de lectura y la mitad inicial del segundo son en realidad las mismas muestras, en el primer caso suavizadas por el fade-out de la ventana de Hanning y en el segundo por su fade-in. Sumando los coeficientes correspondientes a las muestras que intersecan (overlap–add) en este proceso de lectura, se obtienen los coeficientes wavelet de las N muestras de interés.

3.7.2. Estructura de cálculo: convolución circular

Como se ha dicho, en un algoritmo frame-to-frame de tamaño N, los coeficientes no pueden obtenerse mediante la aplicación directa de la Ecuación (1.58). Los problemas que aparecen en la frontera entre frames nacen del tamaño temporal de los filtros pasobanda utilizados (cuya FFT es necesaria de cara a obtener los coeficientes wavelet). De este modo, se hace necesario ajustar el número de muestras de trabajo con aquellas que la FFT necesita para arrojar resultados suficientemente suaves sobre cada filtro, atendiendo a un proceso de convolución circular.

En efecto, de no calcularse la FFT sobre filtros *completos*, al producirse los cambios de frame, los coeficientes obtenidos no tienen por qué ajustarse suavemente en la frontera. Los

errores que se comenten suelen ser numéricamente pequeños, pero tienden a crecer paulatinamente en frecuencias tanto más bajas (produciendo en cada caso un pequeño salto que, acumulado, genera un espurio de alta frecuencia cada N muestras, perfectamente audible).

Para evitar este problema, se ha recurrido al uso de la convolución circular de la FFT unida al mencionado proceso de overlap-add. Se trata, básicamente, de trabajar con el efecto de un filtro completo en t, incluso si los datos que se busca obtener son de un tamaño mucho menor. El proceso detallado aparece representado en la Figura 3.7, que incluye implícitamente la información reflejada en la anterior Figura 3.6.

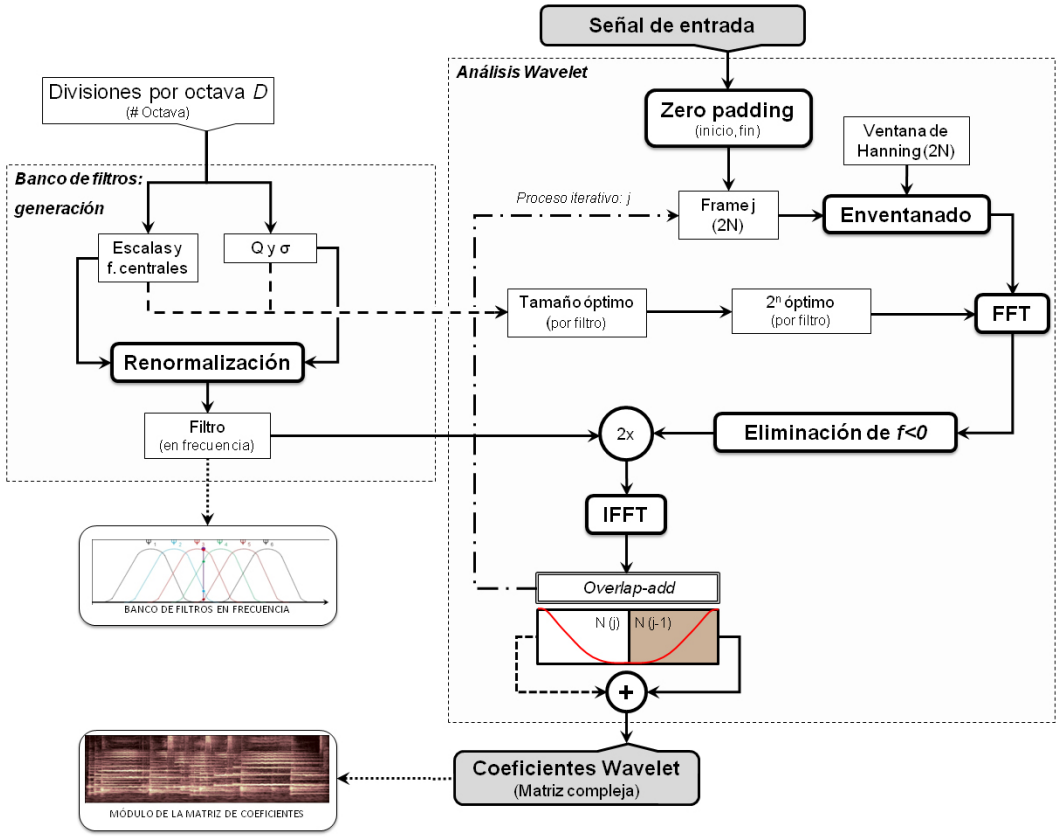


Figura 3.7: *Cálculo de los coeficientes wavelet: diagrama de bloques de la convolución circular.*

En cada análisis se especifica un *vector* de divisiones por octava D para cada octava del análisis. Obviamente, si D es un escalar, la situación es exactamente la detallada en secciones precedentes. Partiendo de este simple vector D, es relativamente sencillo completar las posiciones en escalas y frecuencias centrales de cada uno de los filtros de análisis, sus respectivos anchos de banda y tamaños, en muestras, procediendo como se ha explicado al

final de la Sección 3.3.3. Evidentemente, tratándose de filtros de Q constante, a frecuencias menores (escalas mayores) los anchos de banda frecuencial σ decrecen paulatinamente, y por lo tanto, los filtros son más anchos en t. En otras palabras, a mayor resolución frecuencial (menor σ), mayor deslocalización temporal (filtros más extensos temporalmente). De esta forma, cada uno de los filtros del banco presenta una duración temporal determinada, en muestras. Sea P_n tal duración. Concretamente, se ha tomado:

$$P_n = round(4\sigma_n f_s k_n) \tag{3.35}$$

donde f_s es a frecuencia de muestreo, k_n el valor de la escala (posición frecuencial del filtro $n-simo$) y σ_n su ancho de banda.

La convolución circular de dos secuencias finitas es equivalente a la convolución lineal de tales secuencias, Ecuación (1.58), seguida por un cierto solapamiento temporal [126] (que se tendrá en cuenta en el proceso de overlap–add citado anteriormente). Para aplicar adecuadamente la convolución circular de dos funciones, de tamaños respectivos P_n y N', la Transformada de Fourier debe calcularse sobre un total de muestras [126] S_n tal que:

$$S_n = P_n + N' - 1 \tag{3.36}$$

En este caso, P_n es el tamaño en muestras del filtro n, y $N' = 2N$ el tamaño en muestras leído $x(t)$. Dado que la transformada de Fourier se calculará empleando los algoritmos rápidos de la DFT (FFT), y estos trabajan con potencias enteras de 2, es evidente que en este caso se trabajará sobre una cantidad S' de muestras igual a la potencia de 2 *inmediatamente superior* al dato marcado por S:

$$S'_n = 2^l > S_n, \quad l \in \mathbb{N} \quad | \quad \left\{ \nexists l' \in \mathbb{N} \quad | \quad 2^{l'} > S_n, l' < l \right\} \tag{3.37}$$

A continuación se aplicará directamente la Ecuación (1.58) a la secuencia de datos así obtenida. En primer lugar se calcula la FFT (tamaño S'_n) de los datos enventanados y adecuadamente extendidos (zero padding) de señal y filtro $n-simo$, multiplicándose el resultado por la FFT de los filtros pasobanda del banco. A continuación, se calcula la IFFT de este producto. Los coeficientes wavelet de interés son las primeras N' muestras de esta función (tercer nivel de la Figura 3.6). Como se ha dicho, se guardará en memoria una parte (N muestras) de estos coeficientes, de cara a efectuar el overlap-add final que culmina el proceso. Obtenida de este modo, en la matriz CWT se minimiza la discontinuidad en el proceso de corte en tramas de la información sonora.

El valor máximo de S'_n está en realidad limitado, lo que provoca que para ciertas señales se puedan apreciar las fronteras entre frames en el espectrograma (véase por ejemplo la Figura I.5 en el Anexo I.d). Estos errores son numéricamente despreciables y no resultan

audibles en la resíntesis.

3.8. Corte de parciales

Ahora que se ha detallado el método de cálculo de los coeficientes wavelet, la estructura de datos del espectrograma y del escalograma y el concepto paralelo de parcial que se va a emplear, resulta de capital importancia puntualizar el proceso de selección de las bandas asociadas a cada componente frecuencial de la señal analizada.

Se han ensayado diferentes técnicas de corte de parciales (cada una de ellas con sus ventajas e inconvenientes), entre las que cabe destacar dos, que se detallan en la Figura 3.8: el corte del escalograma por mínimos y la acumulación de parciales de baja energía dentro de zonas de influencia de parciales de mayor contenido de información (zonas de influencia). La elección de la técnica de asignación de bandas puede ser importante a la hora de utilizar el algoritmo CWAS en diferentes aplicaciones. Sin embargo, el resultado final de ambas técnicas es asignar al parcial $i-simo$ detectado (pico $i-simo$ del escalograma), la terna:

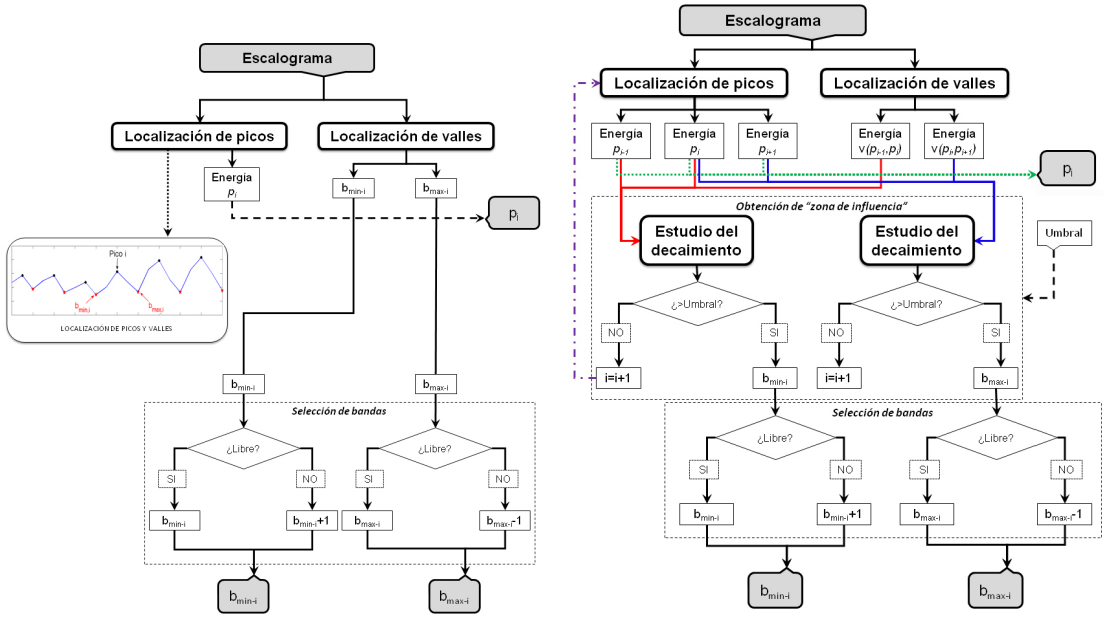
$$(b_{min_i}, b_{pico_i}, b_{max_i}) \equiv (f_{max_i}, f_{pico_i}, f_{min_i}) \tag{3.38}$$

en la cual, la posición del pico (ya sea en bandas o en frecuencias) será utilizada para el proceso de seguimiento de parciales, y los límites mínimo y máximo, además, para construir el parcial en sí.

3.8.1. Corte por mínimos

Como se ha dicho, cada uno de los picos del escalograma wavelet está relacionado con un candidato a parcial. Por el Teorema de Weierstrass, entre dos de tales picos hay un punto en el que el escalograma pasa por un mínimo. Quizá la selección más intuitiva posible de parciales y bandas es que cada pico del escalograma (parcial detectado) quede marcado por la posición de sus mínimos adyacentes superior e inferior. En la Figura 3.9 aparece el escalograma de una señal de guitarra ejecutando una nota $E4$. Con puntos negros se han marcado los picos del escalograma, y con estrellas rojas la posición de los mínimos.

Es evidente que, tomando dos picos consecutivos del escalograma, la banda máxima del pico situado a mayor frecuencia coincide con la banda mínima del siguiente. Por lo tanto, es necesario que uno de los dos parciales "renuncie" a esa banda. De lo contrario, esta será contabilizada dos veces, produciendo un error acumulativo no despreciable. En este caso, la asignación final de bandas se hace por prioridad energética, siguiendo el esquema que aparece en la parte inferior de las Figuras 3.8(a) y 3.8(b), de modo que un parcial determinado quedará delimitado por las bandas b_{min_i} a b_{max_i} más próximas a los mínimos de entre aquellas que no hayan sido elegidas anteriormente. El resultado queda patente en



(a) *Algoritmo de corte por mínimos.* (b) *Algoritmo de corte por acumulación.*

Figura 3.8: *Asignación de bandas: (a) corte directo del escalograma por mínimos, (b) acumulación de parciales por zonas de influencia (valor del umbral de decaimiento, $-3dB$). En ambas figuras, la salida es la terna compuesta por los números p_i (posición del pico), $b_{min,i}$ y $b_{max,i}$ (bandas frecuenciales mínima y máxima asignadas al pico).*

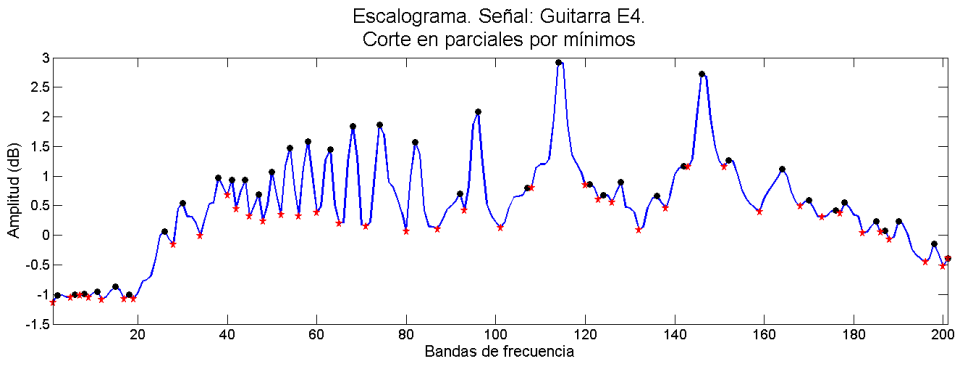


Figura 3.9: *Escalograma de una señal de guitarra. Los picos (puntos negros) son los candidatos iniciales a parciales. Con estrellas rojas, posición inicial de bandas de corte.*

la Figura 3.10, donde las bandas mínimas finales aparecen marcadas con estrellas de dos

tonos. Las desplazadas respecto de la posición real de un mínimo, se han visto afectadas por esta asignación energética.

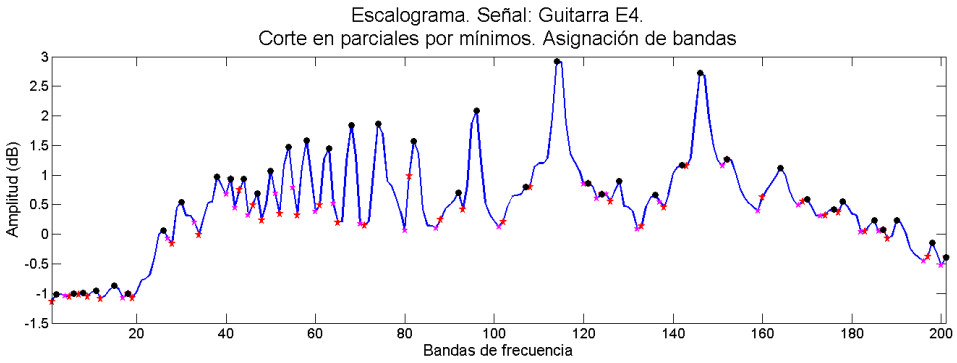


Figura 3.10: *Escalograma de una señal de guitarra. Los parciales (máximos marcados con puntos negros) y sus bandas de corte (estrellas en dos tonos) mínima y máxima relativas.*

3.8.2. Corte en zonas de influencia

Sin embargo, este corte en mínimos presenta un efecto secundario en ocasiones poco deseable: una tendencia a la segmentación, a considerar como parciales lo que puede ser simple ruido de fondo, o información de baja energía (este efecto es especialmente notable en la zona de alta frecuencia). En tales casos, conviene llevar a cabo un corte en bandas ligeramente distinto, reflejado en la Figura 3.8(b).

Partiendo de la misma distribución de "picos" y "valles" del caso anterior, no todos los picos del escalograma se considerarán marcadores finales de parcial, sino sólo aquellos separados por un mínimo que presente un decaimiento suficiente respecto a sus máximos adyacentes. En este caso, el parámetro de decaimiento para que un mínimo sirva como límite de un parcial ha sido tomado a $-3dB$. Es decir, dos picos consecutivos se corresponden a sendos parciales sí y solo sí el mínimo que los separa está al menos $-3dB$ por debajo del *menor* de los picos. Caso contrario, el pico y el mínimo quedan anulados como marcadores, y el candidato a parcial, con sus respectivas bandas, integrado en la zona de influencia del pico de mayor energía.

En la Figura 3.11 aparece reflejado el resultado de la aplicación de esta técnica. Se trata de la misma señal de guitarra de las figuras anteriores. Se puede observar como algunos de los considerados parciales en el caso anterior han sido integrados en estructuras de mayor contenido energético. Por añadido, la asignación final de bandas sufrirá del mismo tipo de desplazamiento respecto a mínimos del que se hablaba en la sección precedente (como

se refleja en la parte baja del bloque algorítmico, Figura 3.8, aunque el resultado de esta corrección no aparece reflejado en la Figura 3.11).

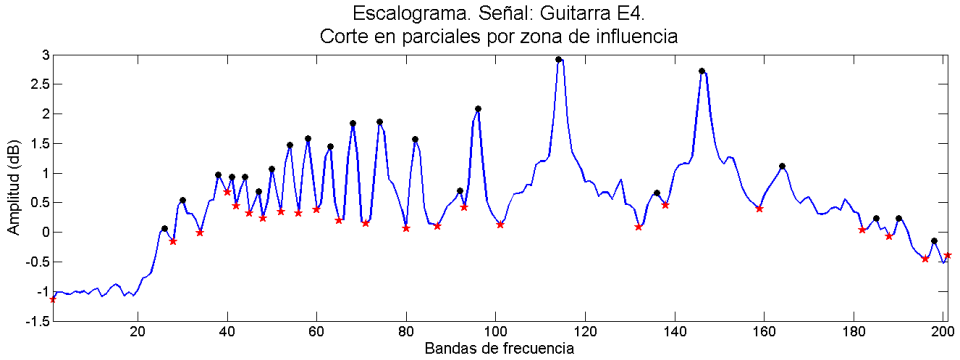


Figura 3.11: *Escalograma de una señal de guitarra. Los parciales (máximos marcados con puntos negros) y sus bandas de corte (estrellas rojas) mínima y máxima relativas, asignadas por zona de influencia.*

El valor de decay escogido, $-3dB$, es suficientemente elevado como para integrar (nunca perder) esa parte superflua de la información, sin que esta tendencia a la fusión resulte excesiva en situaciones en las que su efecto resulta *a priori* menos recomendable, como por ejemplo en el caso de dos parciales que baten.

3.9. Técnicas de Seguimiento (Tracking) de Parciales

Una vez adecuadamente introducida la familia de filtros pasobanda y sus características, indicado el procedimiento de renormalización previo a la resíntesis de la señal o al post-procesamiento de la misma y detallado el procedimiento de asignación de bandas a los parciales detectados, es evidente que el siguiente punto determinante del algoritmo CWAS es el seguimiento (o *tracking*) de tales parciales. En la literatura se pueden encontrar con facilidad diversas técnicas de tracking de parciales imbuidas el contexto del análisis sinusoidal. Un seguidor de parciales eficaz debe ser capaz de extraer constantemente la información adecuada de cara a vincular las componentes espectrales entre marcos sucesivos. Para lograr esta tarea, mayoría de los métodos de seguimiento utilizan la heurística, como la diferencia frecuencial, de amplitud y de fase entre las diferentes componentes localizadas. [52, 116, 150] Una de las técnicas más representativas es la de predicción lineal [104, 105] en la que se emplea la evolución pasada de la información temporal y frecuencial de un parcial para predecir su posible localización futura.

A la largo del desarrollo del algoritmo, se han probado diversas técnicas y efectuado

múltiples modificaciones de las mismas, con el objetivo de encontrar una técnica de tracking óptima para la mayoría de las aplicaciones. En el caso presente, las técnicas que se van a presentar están relacionadas directamente con el seguimiento de picos y zonas de influencia entre escalogramas parciales consecutivos de la señal. En concreto se presentan tres opciones muy distintas entre sí: análisis en bloque, seguimiento punto por punto y tracking frame-to-frame.

Un algoritmo genérico de tracking de parciales aparece en la Figura 3.12. Las técnicas nombradas, que se detallarán en los apartados siguientes, están incluidas dentro de esta figura. El tamaño del frame de señal $(2^n - 1)$ se fija por medio del parámetro n. El tamaño de segmentación del frame (2^m), por medio de m. Si L es el tamaño total de la señal a analizar, se valoran estas tres posibilidades: En primer lugar, $L = n = m$ (donde n y m pueden tener valores no enteros). En tal caso, no cabe hablar de tracking de parciales puesto que la señal entera se analiza en conjunto. Sin embargo, dado que los primeros pasos del algoritmo CWAS se llevaron a cabo aplicando esta técnica, será tratada en primer lugar (Sección 3.9.1). En el otro extremo aparece el tracking punto por punto (Sección 3.9.2). Aquí, independientemente del valor de L y n, se toma $m = 1$. Por último se detallará el análisis trama a trama, en el cual se cumple en general que $L > n > m$ (Sección 3.9.3).

3.9.1. Ejecución en un solo paso

Como se ha dicho, la ejecución en un solo paso no requiere seguimiento de parciales. Sin embargo, dado que es el método original de análisis de la señal, éste es el momento más oportuno para detallarlo mínimamente. En este caso se parte de la información contenida en el escalograma total de la señal. A partir de ésta, se realiza el corte y la asignación de bandas. Las máscaras de cada parcial serán rectangulares, localizadas en las bandas asignadas en cada caso, y de tamaño igual a la duración de la señal en el eje temporal.

3.9.1.1. Resultados

Este tipo de análisis ha sido ejecutado con cierta regularidad, sobre todo en las primeras versiones del algoritmo CWAS y en aplicaciones sencillas. Se pueden encontrar varios ejemplos de señales sintetizadas mediante la aplicación de esta técnica a lo largo de los Capítulos 4 y 5 y en todos los Anexos.

3.9.1.2. Limitaciones

Las limitaciones de esta técnica resultan evidentes. Ante una señal que muestre una excursión frecuencial elevada (por ejemplo el chirp lineal presentado en la próxima Sección), el escalograma total proporcionaría una asignación de bandas que otorgaría una abertura frecuencial enorme a este parcial. Si existe otra componente, (por ejemplo un tono puro)

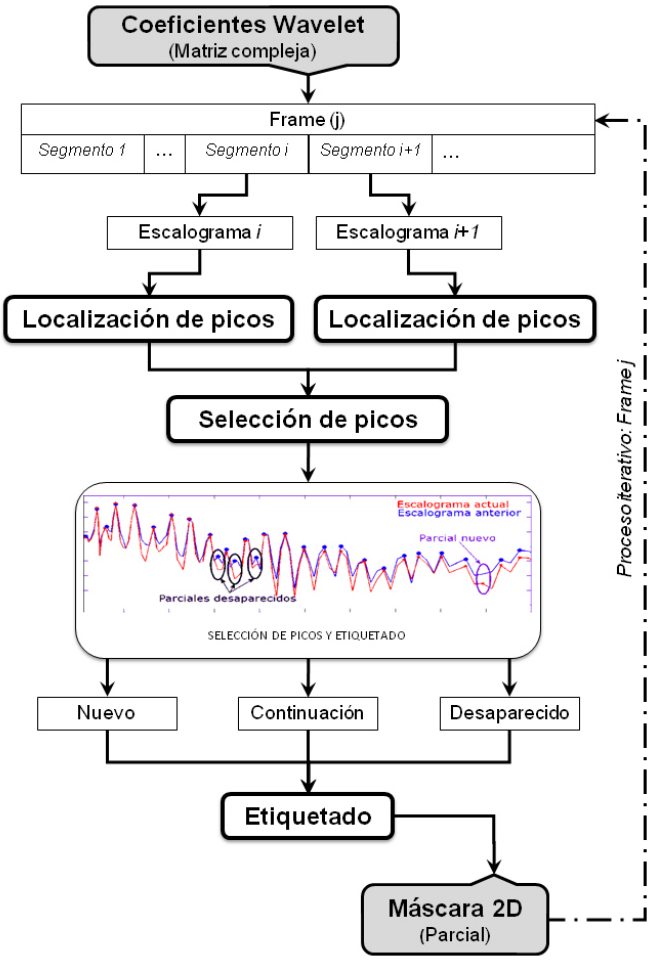


Figura 3.12: *Diagrama esquemático del algoritmo genérico de tracking de parciales.*

y ambas se cortan, resulta muy complicado distinguir qué información se corresponde con qué componente. Afortunadamente, esta situación no suele darse en grabaciones de sonidos reales (sobre todo, sonidos musicales), de modo que la técnica puede aplicarse con cierto éxito incluso en aplicaciones ciertamente complejas como se verá más adelante en la Capítulo 4.

3.9.2. Seguimiento punto por punto

Dado que se dispone de la adecuadamente precisa información frecuencial de la señal en cada instante de tiempo, la más obvia elección en cuanto al tracking de parciales es el

seguimiento *punto por punto*.

En este caso, se efectuará el corte del escalograma instantáneo de la señal (es decir, el resultado del filtrado pasobanda de la misma en cada instante de tiempo). A continuación de entre dos escalogramas consecutivos, se elige el pico del segundo escalograma más próximo a cada uno de los picos del primero, efectuándose de este modo el tracking. El proceso de elección será detallado en la Sección 3.9.3, puesto que el caso actual fue puesto a prueba con un conjunto de señales monocomponente, de modo que la casuística queda definitivamente muy constreñida.

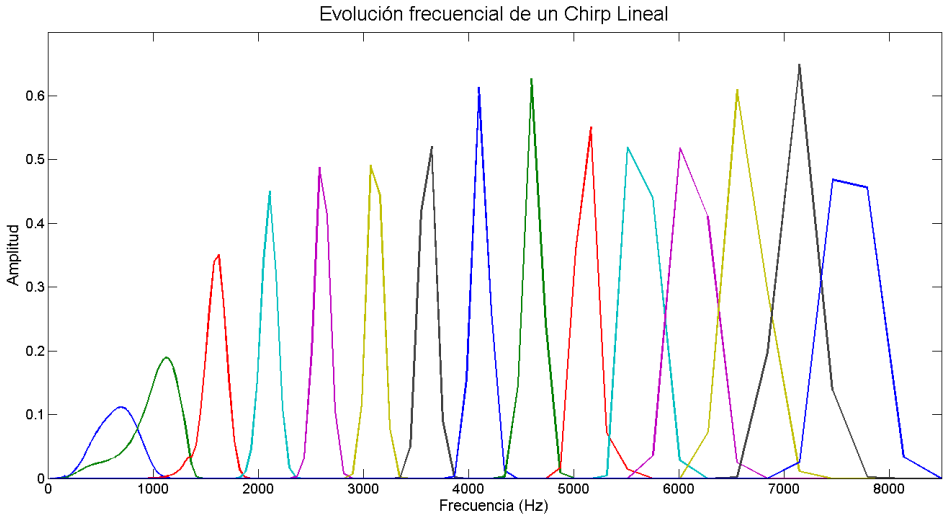


Figura 3.13: *Evolución frecuencial de un Chirp Lineal que varía de 100Hz a 8kHz en 0.5 segundos. La frecuencia de muestreo es f_s =44.1kHz.*

La trayectoria en el semiplano T–F de cada uno de los picos detectados en el escalograma, el esqueleto de la transformada, se empleará más adelante para reconstruir el parcial y, con él, la propia señal.

La evolución temporal del escalograma entre dos instantes de tiempo consecutivos se manifiesta en este caso de forma tan poco aparente, que no resulta práctico mostrarla en una hipotética figura equivalente a la que se presenta en la Sección 3.9.3, (Figura 3.15). Sin embargo, en la Figura 3.13 se ha representado el escalograma de un *Chirp Lineal monocromático*, es decir, una señal (generada sintéticamente) con una sola componente que obedece a la expresión dada por la Ecuación 2.67, en este caso con α=7900Hz, β=100Hz y γ=0. Los escalogramas (en diferentes colores, en la figura) están tomados cada 1400 muestras, lo que explica el valor de 600Hz representado por el primer pico (izquierda de

la figura). La disposición equidistante de los picos es lo que permite calificar como *lineal* al comportamiento frecuencial de esta señal. La duración de $x(t)$ es para este caso de 0.5 segundos, con una frecuencia de muestreo f_s= 44.1kHz y una envolvente de valor máximo 0.99 y 25 milisegundos de suavizado en la entrada y salida de la misma (*fade-in* y *fade-out*, respectivamente). Esta misma señal será utilizada más adelante con otros propósitos.

3.9.2.1. Resultados

Resultados del análisis de señales mediante un tracking de parciales punto por punto aparecen en las Secciones II.d.1 y II.d.2. Al asignar a cada parcial sus bandas asociadas con precisión muestra a muestra, la localización temporal del mismo es la más precisa que se puede obtener mediante el algoritmo CWAS. Esto permite, en casos muy sencillos, extraer información de parciales que están batiendo, pudiéndose incluso superar la limitación del análisis en bandas (Sección II.d.1), si bien el único caso que se ha estudiado en detalle es el de dos parciales unidos. Por otro lado, al asignar de forma tan determinista las bandas de cada parcial, se puede realizar un filtrado de la información en entornos ruidosos. Estos resultados (Sección II.d.2) llegan a limpiar considerablemente la señal incluso en entornos de 0 dB de relación señal a ruido (es decir, un nivel de ruido blanco de energía equiparable a la de la propia señal). Este resultado, sin duda notable, se ha conseguido mediante el análisis de señales sintéticas monocomponente. La técnica resultaría algo más complicada y menos efectiva en el caso de señales reales.

3.9.2.2. Limitaciones

Las limitaciones del tracking punto por punto resultan evidentes. En primer lugar, la toma de decisión sobre qué pico continúa naturalmente a cada uno de los existentes en un instante de tiempo determinado, qué picos son de nueva aparición y qué picos no tienen continuidad, y la ejecución de todas estas operaciones para todas y cada una de las muestras de la señal, ralentiza enormemente la velocidad de cómputo del algoritmo. Este problema se solventa reduciendo el número de operaciones a ejecutar. La pérdida de precisión en la recogida de información frecuencial (otra cuestión de vital importancia), es evidentemente descartable de inmediato, por lo que la única opción es reducir el número de operaciones de seguimiento, es decir, trabajar cada cierto número de muestras en lugar de en cada muestra.

Un problema menor es la introducción de artefactos (grillos) en los parciales sintéticos, debido a la variabilidad de la información recogida. Este problema también se reducirá (si bien no desaparecerá del todo) trabajando por frames y no muestra a muestra.

3.9.3. Seguimiento trama a trama

La decisión final tomada para el algoritmo CWAS más básico ha sido el análisis *frame-to-frame*. Como se ha concluido en la sección precedente, esto reduce el número de operaciones a efectuar, con la consiguiente disminución del tiempo de procesado. Además, permite leer la señal en tramas de cierto tamaño, reduciendo por lo tanto el volumen de datos a estudiar en un momento dado, lo cual reduce el nivel de acceso a disco duro, permitiendo así mismo ganar en velocidad de proceso y, de la misma forma, analizar señales de mayor duración. Frames de análisis más grandes implican mayores ganancias, pero a su vez se tiende a la situación inicial (análisis en un sólo paso), con sus correspondientes inconvenientes. Por lo tanto, es necesario encontrar un tamaño de frame adecuado que respete la precisión del resultado sin resultar prohibitivo en tiempo de proceso.

3.9.3.1. Elección del tamaño de trama

Este es el motivo por el que se va a trabajar por defecto en un algoritmo frame-to-frame *anidado*, como el que aparece representado en la Figura 3.14. Considérese una señal

$N=2^n-1$				$N=2^n-1$				…	$N_L=L-(J-1)N$			
j=1				j=2				…	j=J			
i=1	i=2	…	i=I	i=1	i=2	…	i=I	…	i=1	i=2	…	i=I'
$M=2^m$	$M=2^m$	…	$M'=2^m-1$	$M=2^m$	$M=2^m$	…	$M'=2^m-1$	…	$M=2^m$	$M=2^m$	…	$M''=N_L-(I'-1)M$

Figura 3.14: *Representación esquemática de los tamaños de frame (j) y segmento (i) del algoritmo frame-to-frame anidado, para n, m y L genéricos.*

de duración total L, en muestras. En lo que resta, salvo indicación expresa de lo contrario, los frames iniciales son de tamaño $n = 12$, (es decir, N =4095 muestras), y los segmentos pequeños de $m = 8$ ($M = 256$ muestras). Esto significa que cada frame de N muestras será procesado a su vez en I (en este caso $I = 16$) partes de M muestras cada una (salvo la última, de $M' = 255$). El último frame de la señal tendrá la duración en muestras la cola final de señal sin analizar, $N_L = L - (J-1)N$, siendo J la duración total en frames de la señal (incluyendo el frame final, irregular). Por lo tanto, en general su tamaño será $N_L < N$, y será analizado en $I' - 1 < I$ segmentos de M muestras, más un segmento final de tamaño $M'' = N_L - (I'-1)M$.

La elección de estos valores no es aleatoria. Por un lado, se reduce el tiempo de procesado respecto al caso muestra a muestra de manera significativa (*grosso modo*, en un factor M). Sin embargo, para reunir información modular y sobre todo frecuencial suficiente para

muchas de las aplicaciones posteriores, se necesitan más muestras. De aquí las $N = 4095$ tomadas como base (que equivalen a 0.0929 segundos si $f_s = 44100$Hz, y a 0.1857 segundos para $f_s = 22050$Hz). De esta forma, cada una o dos décimas de segundo se acumula suficiente información de módulo y frecuencia instantáneos como para, por ejemplo, poder efectuar un análisis frecuencial o una una separación de fuentes con cierto margen de garantía.

3.9.3.2. Procedimiento

El procedimiento a través del cual se lleva a cabo el seguimiento de las componentes detectadas es muy simple, y aparece resumido en el diagrama de bloques de la Figura 3.12. Se parte de la información contenida en dos segmentos consecutivos (i, $i+1$) dentro de un frame determinado (j). El algoritmo de tracking compara la posición de los picos dentro del segmento actual ($i+1$) con los datos correspondientes al segmento anterior (i). Para cada pico del escalograma i caben tres posibilidades:

1. Existe una correspondencia biunívoca este pico y algún otro del segmento $i+1$.
   - El sistema asigna la misma etiqueta a ambos picos.
2. No existe un pico en $i+1$ correspondiente al pico en cuestión del segmento i.
   - El sistema cierra la información del parcial correspondiente al pico desaparecido.
3. Existe algún pico de $i+1$ que no se corresponde con ninguno del escalograma i.
   - El sistema asigna una etiqueta (libre) a estos parciales de nueva aparición.

Todas estas posibilidades aparecen reflejadas en la Figura 3.15. Se trata de una ampliación de la zona frecuencial comprendida entre los 20Hz y los 2kHz de una señal de guitarra ejecutando una nota $E2$, de fundamental $f_0 = 81.8260$Hz. El escalograma $i = 5$ aparece en azul, con sus picos correspondientes (parciales detectados), en dB, marcados con círculos del mismo color. El escalograma del segmento siguiente, $i+1 = 6$ aparece en rojo, con sus picos señalados por estrellas. Como se puede apreciar en la figura, la mayoría de los parciales del escalograma inicial tienen un compañero fácilmente asignable en el escalograma siguiente. Sin embargo, algunos parciales (picos) no tienen continuidad (marcados con elipses negras en la Figura), mientras que otros son nuevos (en morado). El algoritmo guardará en memoria la posición de los picos del segundo escalograma $i+1$ junto con sus correspondientes bandas de corte asociadas. Los datos irán a parar a una máscara pre-existente, si se corresponden al caso 1; a una máscara de nueva creación en el caso 3 (adecuadamente rellenada de ceros hasta la posición marcada por j e $i+1$); o guardará ceros en la máscara adecuada frente a la situación contemplada en el caso 2.

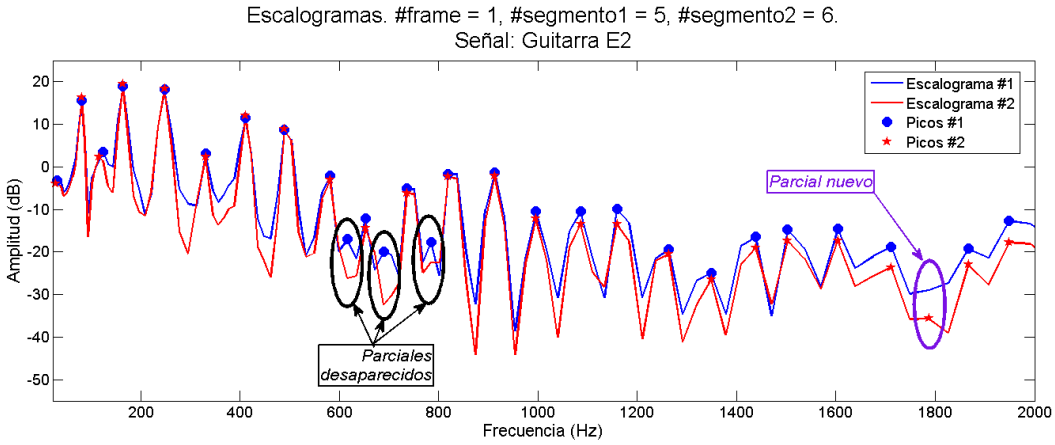


Figura 3.15: *Ejemplo de tracking de parciales. Señal: Guitarra E2 ($f_0 = 81.8260Hz$). Número de frame: $j = 1$. Número de segmento del escalograma anterior: $i = 5$. Número de segmento del escalograma actual: $i + 1 = 6$.*

De esta forma, rellenando los datos correspondientes al segmento de estudio dentro del frame abierto en la posición adecuada de cada máscara bidimensional etiquetada, se van completando la información relacionada con cada uno de los distintos parciales detectados en la señal.

3.9.3.3. Resultados

El análisis frame-to-frame anidado ha sido utilizado en las aplicaciones presentadas en las Secciones 3.10, 5.5.1, 5.5.2, 4.6, III.c (b) y (c), II.d.3 y 5.5. Por lo tanto, será considerado en adelante como la base algorítmica de trabajo. Los errores en la resíntesis obtenidos con esta técnica son despreciables tanto numérica como acústicamente. El tiempo de procesado (dependiendo de las características de la máquina, del sistema operativo y de la versión de Matlab® instalada) oscila en torno a 15× (procesador Intel® $\text{Core}_{(TM)}$ i7 @ 2.67GHz, 12GB de RAM, sistema operativo Windows 7 64bits, Matlab® 7.8.0.347) es decir, unos 15 segundos por cada segundo de duración de la señal a analizar (este tiempo incluye el cálculo de los coeficientes wavelet y el post procesado de la información). Variaciones en la versión de Matlab e incluso en el sistema operativo (Windows 7 64bits) alteran en gran medida estos tiempos de proceso.

3.9.3.4. Limitaciones

Las limitaciones de este método son básicamente dos. En primer lugar, el tiempo de procesado, que descarta de forma inmediata las aplicaciones en tiempo real del algoritmo. Por otro lado, la segmentación en la información de los parciales detectados puede resultar un lastre (dependiendo de la aplicación). Los parciales sintéticos de menor energía tienen cierta tendencia a presentar algún artefacto (si bien ni la señal sintética ni otras señales obtenidas a partir de estos parciales muestran defectos evidentes). Como renunciar a la información contenida en estos parciales supone añadir un error nuevo a los resultados de resíntesis, en general todos ellos serán tenidos en cuenta en el proceso de síntesis aditiva. Parte de la segmentación puede evitarse fusionando los parciales de corta duración y/o baja energía con otras componentes más importantes con las que compartan frontera frecuencial.

3.10. Sonidos sintéticos

La primera y más directa aplicación del algoritmo detallado hasta ahora, es la síntesis de sonidos: el análisis de las señales de audio sin otro objetivo que poder generar un sonido sintético indistinguible del original. Para ello se ejecuta el algoritmo CWAS, recogiéndose la información de la máscara correspondiente a cada parcial detectado de la señal, de forma individual. Esta máscara es desplegada sobre la matriz CWT compleja (cuyo módulo es el espectrograma wavelet), extrayéndose los coeficientes complejos involucrados en ella, los cuales son adecuadamente sumados en el eje frecuencial para obtener la función compleja correspondiente al parcial a través de la Ecuación (3.18). La parte real de esta información es la contribución del parcial a la señal total, la cual se generará a través de la síntesis aditiva de tales parciales, Ecuación (3.24). De este modo se puede sintetizar una señal de salida de características extremadamente similares a la original. Esta aplicación inicial es la que ha dado nombre al algoritmo: síntesis Aditiva por Wavelets Complejas (Complex Wavelet Additive Synthesis), CWAS.

La síntesis aditiva de los parciales complejos obtenidos en el análisis permite obtener un modelo de la señal coherente a la vez, como se verá en el próximo capítulo, en tiempo y en frecuencia. De este modo se hace posible calcular el error temporal $e(t)$ cometido, *restando muestra por muestra la señal original y la sintética*. El orden de magnitud promedio de los errores temporales así calculados está situado alrededor de -50dB. Para que tales errores resulten visibles, suele ser necesario amplificarlos.

En la Figura 3.16 se ha representado gráficamente el resultado de la síntesis por el algoritmo CWAS de la misma señal de guitarra, ejecutando una nota de f_0=329.88Hz ($E4$), utilizada para los escalogramas de las Figuras 3.9, 3.10 y 3.11. La señal original ha sido muestreada con f_s = 44.1kHz. En ella se han localizado y seguido hasta 60 parciales dife-

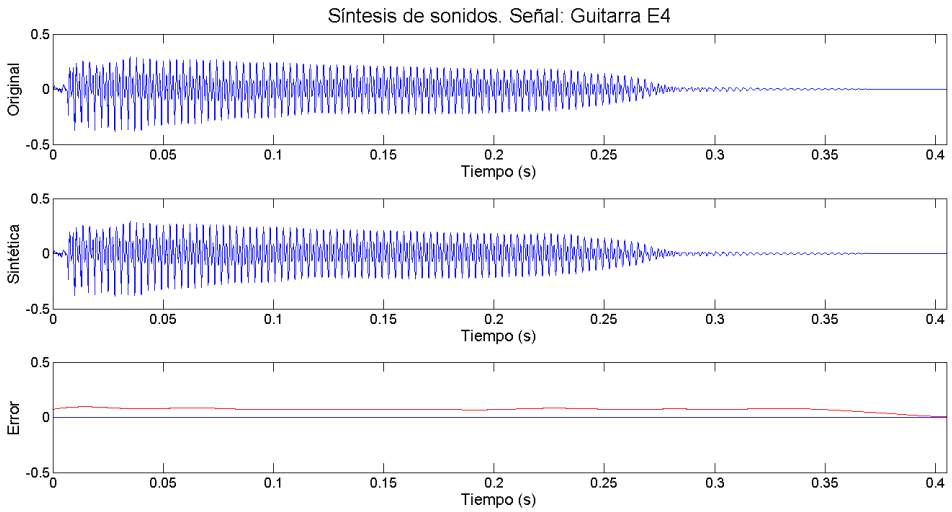


Figura 3.16: *Síntesis de una guitarra ejecutando una E4. Arriba: forma de onda original. En el centro: forma de onda de la señal sintética. Abajo: en trazo azul, el error cometido. En trazo rojo, el error cometido amplificado en un factor 100.*

rentes.

El error cometido $e(t)$ aparece representado en la gráfica inferior de la Figura 3.16; en trazo rojo su valor multiplicado por 100. Evidentemente, la señal original y la sintética resultan muy complicadas de distinguir acústicamente (véase Anexo I).

Esta y otras aplicaciones serán estudiadas con más detalle en los Capítulos 4 y 5.

3.11. Conclusiones y contribuciones

En este Capítulo se ha procedido a la explicación exhaustiva de los bloques principales que componen el algoritmo CWAS. Partiendo de la generación del banco de filtros pasobanda utilizado, y tras redefinir el concepto de parcial, se llega a un modelo de renormalización ciega diferente al propuesto en la Sección 2.4.1.1. Tras la obtención de la matriz de coeficientes, se explican y ensayan distintas técnicas de corte en bandas y de tracking de parciales. Con la información generada hasta este punto, el algoritmo CWAS está en condiciones de ser puesto a prueba en diferentes aplicaciones, la más directa de las cuales es la resíntesis de señales (Capítulo 5). No obstante, en el siguiente Capítulo nos centraremos en la más completa de ellas, la separación ciega de fuentes monaurales de sonido.

En cuanto a las contribuciones presentadas en este Capítulo, son:

1. Obtención de un banco de filtros flexible partiendo de la wavelet de Morlet modificada (ver Sección 2.2):
   - De base diádica.
   - Parámetro de control: Divisiones por octava (D).
2. Redefinición del concepto de *parcial*, basado en *zonas de influencia*.
3. Obtención del parámetro de *renormalización efectiva*, basado en el concepto de *sobrepeso*.
4. Evaluación de diferentes técnicas de asignación de bandas a parciales.
   - Corte por mínimos.
   - Corte por decaimiento (zonas de influencia).
5. Evaluación de diferentes técnicas de seguimiento o *tracking* de parciales.
   - Ejecución en un solo paso.
   - Tracking punto a punto.
   - Tracking frame to frame.

Como se puede comprobar, algunas de las contribuciones presentadas son de creación exclusiva para el algoritmo CWAS, mientras que otras se tratan de la adecuación de técnicas existentes a la herramienta propuesta.

Capítulo 4

Separación ciega de notas en fuentes de audio monaurales

Índice

"El modo de dar una vez
en el clavo es dar cien veces
en la herradura".
Miguel de Unamuno (1864-1936).
Ensayista, novelista, poeta
y periodista español.

La aplicación más completa que se ha desarrollado en esta Tesis es sin duda la separación ciega de fuentes de audio, objeto del presente Capítulo. Aunque también se ha abordado una primera aproximación a la separación de fuentes en mezclas estereofónicas, en el presente capítulo se detallará exclusivamente una técnica de separación monaural basada en la reconstrucción de amplitud y fase de los parciales superpuestos. Para detallar esta técnica se hace necesario exponer además los algoritmos de búsqueda de onsets y de estimación de frecuencias fundamentales que se han desarrollado. En el estado actual del procedimiento propuesto, se puede abordar la separación de notas musicales de dos o tres fuentes diferentes (correspondientes al mismo o a diferentes instrumentos musicales), con resultados acústicos de gran calidad. Se puede consultar información adicional en el Anexo III.

4.1. Introducción

El algoritmo de separación ciega de fuentes de audio monaurales que se presentará en el presente Capítulo ha evolucionado en tres fases:

1. En la primera aproximación al problema [21], de carácter tal vez más general, se emplean los tiempos de onset para clasificar las fuentes presentes en familias, y de este modo reducir el costoso tratamiento estadístico de la información de partida. Para más detalles, consúltese el Anexo III.b.

2. El segundo estadio [55] se centra ya en señales armónicas. Se trata de un algoritmo frame-to-frame en el que la información de alta frecuencia es ignorada, lo cual aumenta la calidad numérica de los resultados respecto al primer método de separación a costa de que las señales separadas pierdan buena parte de su color característico, sonando "huecas". Esta técnica está detallada en el Anexo III.c.

3. La técnica final [18] pasa por intentar estimar la contribución a los parciales mezcla sin emplear en la síntesis la información *real* procedente de tales parciales. En su lugar, se genera información completamente nueva, partiendo de parciales aislados. Como consecuencia de esto, tanto la calidad numérica de la separación como la acústica de las señales separadas resulta especialmente elevada.

Hay que dejar claro que ninguna de estas técnicas se encuentra en un estado lo suficientemente avanzado como para poder separar temas musicales. Sin embargo, se ha tratado de dejar al algoritmo CWAS bien dirigido para que los avances en esta linea puedan ser razonablemente rápidos.

Como se puede ver en el Anexo III, los dos primeros métodos están basados en *distancias* y no abordan la separación de parciales superpuestos. La última técnica será desarrollada a lo largo de las siguientes Secciones y es con diferencia la más completa que se ha implementado. Como se ha avanzado, es capaz de separar notas musicales de dos o tres fuentes diferentes (interpretadas por el mismo o por diferentes instrumentos) con resultados acústicos y numéricos de gran calidad.

4.2. La Separación Ciega de Fuentes: un problema complejo

Dada su complejidad y actual grado de actividad investigadora internacional, la *separación ciega de fuentes de audio* merece un capítulo propio. Se trata de un tema que ha recibido una atención investigadora creciente en los últimos años. La búsqueda más superficial de bibliografía en esta rama del audio proporciona centenares de artículos de investigación, la mayoría de los cuales ha aparecido en la última década.

Las técnicas de BASS intentan recuperar las señales fuente partiendo de una entrada $x(t)$ mezclada, cuando el proceso de mezcla resulta, por lo demás, desconocido. El término "ciega" significa que la información previa (concerniente a la señal) necesaria para llevar a cabo la separación es muy poca (idealmente ninguna, aunque en general necesitaremos hacer una serie de suposiciones que tienden a simplificar el problema).

En el caso más general, representado en la Figura 4.1, la separación contará con N fuentes mezcladas y M mezclas diferentes recogidas por otros tantos micrófonos. Al igual que en los sistemas matemáticos de ecuaciones, el número de canales de mezcla define cada situación particular, y para cada una de éstas la literatura provee de varios métodos de separación. Probablemente el caso más estudiado e interesante es el indeterminado, en el cual $N > M$ (aquí la analogía con las matemáticas falla: un menor número de micrófonos no sólo no conlleva que la separación quede irresoluta; ni siquiera implica necesariamente que se obtengan peores resultados). Por ejemplo, en la separación estéreo (a través del algoritmo DUET [135] y posteriores evoluciones de las técnicas de enmascaramiento Tiempo–Frecuencia [46, 47, 117]), el retardo y la atenuación relativos entre los canales izquierdo y

derecho puede ser utilizada para discriminar las fuentes presentes, así como para llevar a cabo algún tipo de análisis de la escena musical [175].

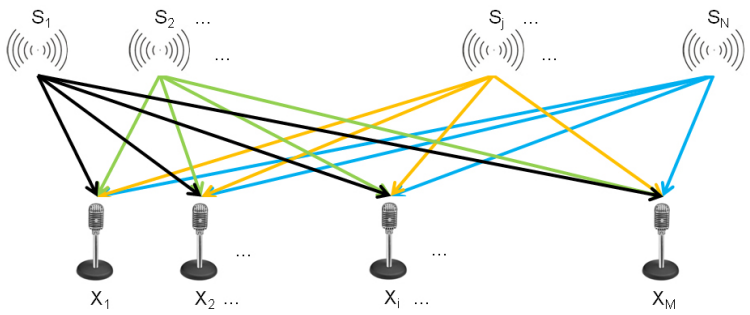


Figura 4.1: *Representación simbólica del problema de la mezcla de N fuentes capturada con M micrófonos.*

La situación es bastante más compleja cuando nos enfrentamos al caso monaural, es decir, cuando $M = 1$. En tal situación no se dispone de más información para llevar a cabo el proceso de separación que la propia señal mezcla. Sin embargo, el oído humano es capaz de discriminar las fuentes presentes en la mezcla incluso en estas circunstancias [34], permitiendo a las personas sordas de un oído prestar más atención a una fuente de sonido determinada por encima de las demás (si bien resulta imposible discernir la dirección del sonido y por lo tanto llevar a cabo el análisis de la escena sonora). La alta indeterminación matemática del problema incrementa enormemente las dificultades de la separación. Por lo tanto, la separación monaural es el desafío más complicado dentro de la separación de fuentes.

4.2.1. Técnicas de BASS: un breve repaso

A lo largo de los años se han desarrollado múltiples técnicas de separación ciega de fuentes en general, alguna de las cuales abarca el caso de la separación monaural. De forma grosera, se pueden dividir en tres grandes categorías: estudios de carácter psicoacústico, técnicas estadísticas y técnicas de modelado sinusoidal.

Los estudios de carácter psicoacústico como el Análisis de la Escena Auditiva por Computador (*Computational Auditory Scene Analysis*, CASA) [37, 170], inspirado en el Análisis de la Escena Auditiva (*Auditory Scene Analysis*, ASA) [34], el cual sintetiza una teoría descriptiva sobre cómo el cerebro procesa la información auditiva de la cloquea para extraer y seguir las señales relevantes en medios ruidosos o interferentes (atención selectiva). La separación ocurre a través de la unión y agrupación de los acontecimientos en el dominio tiempo-frecuencia [43], lo cual sugiere que la coherencia espectral y temporal entre las

fuentes puede ser muy importante a la hora de discriminarlas.

Dentro de las técnicas estadísticas, el Análisis de Componentes Independientes (*Independent Component Analysis*, ICA) [5, 40] asume cierta independencia estadística entre las fuentes, mientras que el Análisis de Subespacios Independientes (*Independent Subspace Analysis*, ISA) [170], extiende las bases de ICA hasta la separación monocanal. La técnica de Descomposiciones Dispersas (*Sparse Decompositions*, SD) [85] asume que una fuente es una suma ponderada de bases de un conjunto incompleto, considerando que la mayoría de estas bases permanecen inactivas durante la mayor parte del tiempo [1], es decir, se presume que sus pesos relativos sean mayoritariamente cero. La Factorización de Matrices No-negativas (*Nonnegative Matrix Factorization*, NMF) [144, 167], intenta encontrar una matriz mezcla (con pesos distribuidos [108, 143]) y una matriz fuente con elementos no negativos tales que el error en la reconstrucción se minimice.

Por último, en las técnicas de modelado sinusoidal se asume que todos los sonidos son una combinación lineal de sinusoides con amplitudes, fases y frecuencias variables en el tiempo. Por lo tanto, la separación requiere de la estimación adecuada de estas variables para cada fuente presente en la mezcla [57, 166, 168], o algún tipo de conocimiento a priori de otras características, como por ejemplo una estimación inicial de los *pitches* individuales [110, 173]. Una de las aplicaciones más importantes de esta técnica es la mejora de la inteligibilidad del habla [84], basada en la separación voz-interferencia de cara a amplificar la primera o atenuar la segunda. La mayoría de los autores que trabajan en el campo de la separación monaural por modelado sinusoidal trabajan con la STFT de cara a analizar la señal mezclada para obtener sus principales componentes sinusoidales. También pueden utilizarse las llamadas representaciones basadas en el auditorio, *Auditory-based representations* [38], en las que el eje frecuencial se deforma siguiendo una escala de frecuencias auditivas, que enfatizan la resolución (mediante técnicas como el ancho de banda igual rectangular y las escalas de Bark) en el rango medio-bajo de frecuencias, donde generalmente se concentra más la energía del sonido.

4.2.2. Parciales aislados y parciales superpuestos

Atendiendo al espectro de una señal mezclada, resulta evidente que existirán zonas espectrales pertenecientes básicamente a una u otra fuente, mientras que otras se corresponderán con zonas donde la información de ambas señales queda superpuesta. Para ilustrar estas posibilidades se ha incluido la Figura 4.2, donde se representa la zona frecuencial entre 20Hz y 3kHz del escalograma de una señal de clarinete (linea azul a trazos) y una flauta (linea roja punteada), así como el escalograma de la señal mezcla (linea negra continua). En la Figura, el eje vertical es logarítmico, para resaltar los datos.

Como se puede observar, es evidente que realmente en todo el espectro aparece si-

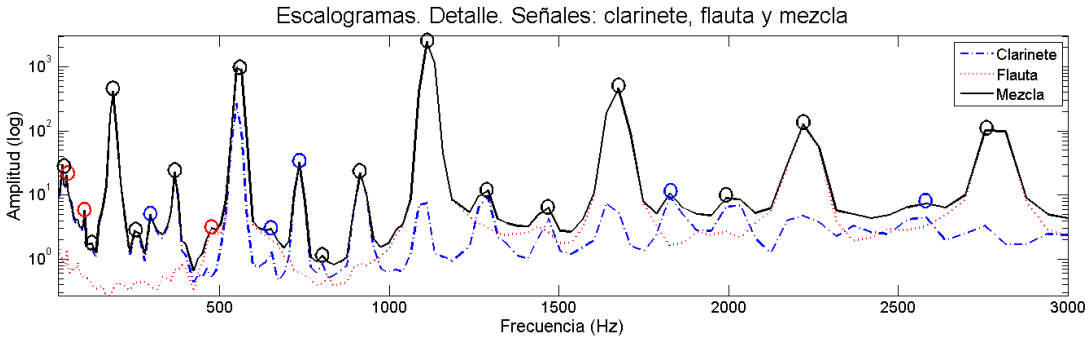


Figura 4.2: *Detalle de los escalogramas correspondientes a la señal de clarinete (en azul), flauta (en rojo) y mezcla (en negro). Se ha resaltado con círculos de colores la posición del máximo relativo a cada parcial, así como si se corresponden mayoritariamente a una u otra fuente o bien son compartidos (círculos azules, rojos y negros, respectivamente).*

multáneamente información correspondiente a ambas señales (y por lo tanto no existen parciales que contengan información *exclusivamente* de una de las fuentes aislada, lo cual garantiza que habrá interferencias en mayor o menor medida). Sin embargo, en algunos casos la inmensa mayor parte de la información recogida en el parcial mezcla (hay que recordar que un parcial es toda la zona comprendida entre mínimos, no simplemente un pico) se corresponde mayoritariamente con información de una u otra fuente, mientras que otros, cuya información se superpone más claramente en el escalograma total, serán parciales superpuestos. En la Figura 4.2, estos tres tipos de parciales se han señalado sobre la señal mezcla con círculos azules, rojos y negros, siguiendo la misma leyenda que los diferentes escalogramas representados.

Un algoritmo de separación ciega debe ser capaz de extrapolar, partiendo únicamente de la mezcla, cuántas fuentes hay presentes en la señal y qué información espectral se corresponde con cada una de ellas. Finalmente, habrá que abordar el tema de la separación de la información interferente.

4.2.3. Separación de parciales superpuestos: estado del arte

La separación de parciales superpuestos en sonidos tonales es uno de los problemas más complicados de resolver (tanto analítica como algorítmicamente) en BASS. De hecho, muchos autores suponen la ortogonalidad o disjunción entre fuentes, de modo que el problema de parciales superpuestos no es tenido en cuenta [168]. La superposición de espectros ha sido estudiada sin embargo durante varias décadas [129], si bien no ha sido hasta los últimos años que se ha producido un incremento significativo en la investigación acerca de este tópico. En tanto en cuanto la información de las regiones superpuestas en el semiplano T–F resulta

poco fiable, varios sistemas recientes intentan utilizar la información de parciales vecinos no superpuestos para de algún modo estimar las contribuciones de cada fuente en las zonas compartidas. Algunos sistemas asumen que la envolvente espectral del sonido de un instrumento es suave [91, 169], y que por lo tanto la amplitud de un armónico superpuesto puede ser inferida a partir de las amplitudes de armónicos no superpuestos de la misma fuente via interpolación [57, 169] o suma ponderada [166]. Sin embargo, esta suavidad espectral se viola a menudo en las grabaciones de instrumentos musicales reales. Una aproximación diferente se conoce como Modulación de Amplitud Común (*Common Amplitude Modulation*, CAM) [110], en la que se asume que las envolventes temporales de armónicos diferentes de la misma fuente tienden a ser similares. Otra opción es la técnica de la Estructura Armónica Media [54] (Average Harmonic Structure, AHS) en la cual, dado el número de fuentes, se crea un modelo de estructura armónica para cada una de ellas, pudiendo emplearse esta información para separar armónicos superpuestos. En [79], los autores proponen un modo de operación alternativo para la estimación de envolventes armónicas, la Similitud de Envolvente Temporal de Armónicos (Harmonic Temporal Envelope Similarity, HTES). En este caso, se emplea la información de los armónicos no superpuestos de las notas de un instrumento (ocurran cuando ocurran dentro de una grabación) para crear un modelo del mismo que puede ser utilizado para obtener las envolventes de los parciales compartidos, permitiendo de este modo la separación incluso de notas completamente superpuestas.

4.3. Parámetros numéricos estándar de calidad

En este trabajo se asumirá que los errores cometidos en la separación tienen tres orígenes diferentes: pueden ser debidos a términos de interferencia entre fuentes, a distorsiones insertadas en la señal separada y a los artefactos propios del algoritmo de separación en sí. El ruido no se ha tenido en cuenta.

Desde este punto de vista, en una mezcla de N fuentes dada por la Ecuación (4.8), sean s_k las señales originales y $\bar{s}_k$ las fuentes separadas. La distorsión total relativa puede definirse como [66, 71, 165]:

$$D_{total} = \frac{\|\bar{s}_k\|^2 - |\langle \bar{s}_k, s_k \rangle|^2}{|\langle \bar{s}_k, s_k \rangle|^2} \tag{4.1}$$

Asumiendo una descomposición ortogonal:

$$\bar{s}_k = \langle \bar{s}_k, s_k \rangle s_k + \varepsilon_{interf} + \varepsilon_{artif} \tag{4.2}$$

donde $\langle \bar{s}_k, s_k \rangle$ es la contribución de la fuente objetivo, ε_{interf} es el término de error debido a la interferencia de las demás fuentes, y ε_{artif} es el término de error debido a los artefactos

generados por el algoritmo de separación.

La distorsión relativa debida a interferencias se puede definir, en este caso, como:

$$D_{interf} = \frac{\|\varepsilon_{interf}\|^2}{|\langle \bar{s}_k, s_k \rangle|^2} \tag{4.3}$$

Por otro lado, la distorsión debida a artefactos sería:

$$D_{artif} = \frac{\|\varepsilon_{artif}\|^2}{\|\langle \bar{s}_k, s_k \rangle s_k + \varepsilon_{interf}\|^2} \tag{4.4}$$

Se han empleado tres parámetros numéricos estándar para probar la calidad final de los resultados de separación. Estos parámetros son la relación señal a interferencia, (*signal-to-interference-ratio*, SIR), la relación señal a distorsión (*signal-to-distortion-ratio*, SDR) y por último la relación señal a artefactos (*signal-to-artifacts-ratio*, SAR):

$$SIR = 10 \log_{10} \left(D_{interf}^{-1} \right) \tag{4.5}$$

$$SDR = 10 \log_{10} \left(D_{total}^{-1} \right) \tag{4.6}$$

y:

$$SAR = 10 \log_{10} \left(D_{artif}^{-1} \right) \tag{4.7}$$

Los valores numéricos de estos parámetros se han obtenido en el entorno de la herramienta de *MATLAB* ® conocida como *BSS_EVAL*, desarrollada por Févotte, Gribonval, y Vincent y distribuida bajo Licencia Pública GNU [66].

4.4. Separación monaural de sonidos musicales

El problema de la separación ciega de fuentes de audio monaurales es, como ya se ha dicho, uno de los más complicados e interesantes dentro de las más activas líneas de investigación a nivel internacional. Se trata de separar un número arbitrario N de fuentes registradas por un único micrófono:

$$x(t) = \sum_{k=1}^{N} s_k(t) \tag{4.8}$$

En esta ecuación, $x(t)$ es la señal mezclada, mientras que $s_k(t)$ son las fuentes originales.

Como se ha dicho, un algoritmo de separación ciega debe ser capaz de segregar, partiendo de $x(t)$, cuantas fuentes $s_k(t)$ hay presentes en la señal y qué información se corresponde con cada una de ellas. Una forma de conseguir esta información consiste en llevar a cabo una es-

timación de las diferentes frecuencias fundamentales presentes en un momento determinado. A partir de ellas pueden generarse los diferentes peines armónicos teóricos, comparándose los resultados con el espectro de $x(t)$ y obteniéndose una primera división de los parciales en *aislados* (pertenecientes a una fuente determinada) y *superpuestos* (pertenecientes a más de una fuente). Empleando alguna de las técnicas explicadas en la Sección 4.2.3 se separarían las contribuciones pertenecientes a cada fuente, generándose de este modo las diferentes señales $s_k(t)$.

El algoritmo de separación propuesto se presenta en la Figura 4.3.

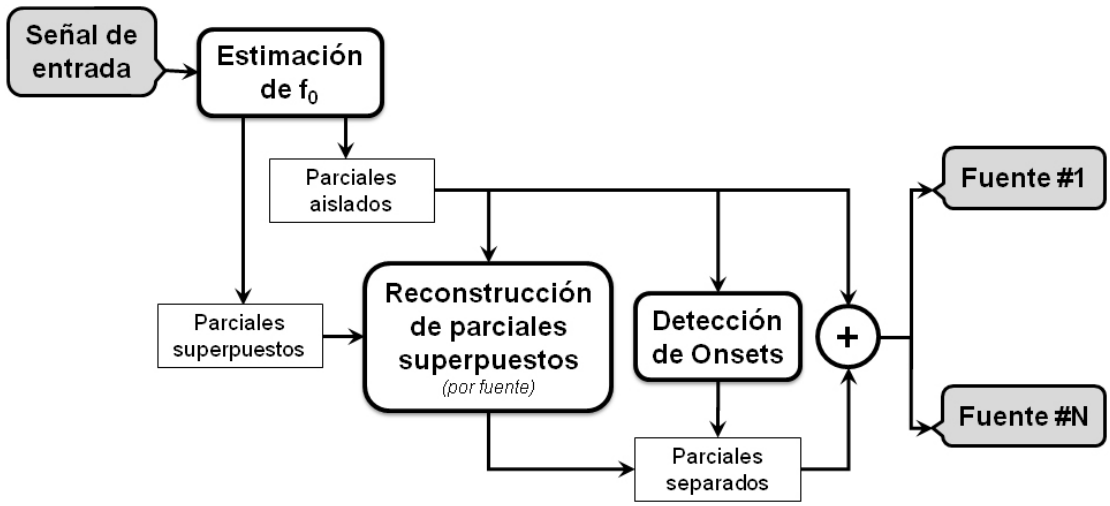


Figura 4.3: *Diagrama de bloques del algoritmo de separación propuesto.*

Partiendo únicamente de la señal mezclada, se procede a la estimación de todas las frecuencias fundamentales presentes en $x(t)$, de modo que se tendrán presentes en principio sólo las fuentes armónicas de la mezcla. A partir de las frecuencias fundamentales se construyen los distintos peines armónicos correspondientes a cada fuente, y los parciales de la señal se dividen en aislados y superpuestos. Empleando una adaptación del principio CAM (el cual supone que las amplitudes de los parciales correspondientes a una misma fuente están altamente correladas) y la aproximación armónica de las fases involucradas, se lleva a cabo la reconstrucción de los parciales superpuestos. Mediante un algoritmo de detección de onsets se pueden eliminar la información anterior y posterior a la ejecución de cada nota (evitándose de este modo la presencia de colas espurias pertenecientes a otras fuentes). Finalmente, la suma de los parciales aislados de cada fuente y la contribución a la misma que pueda haber en los diferentes parciales superpuestos, se genera cada una de las señales $s_k(t)$ correspondientes a las fuentes presentes.

Antes de abordar de forma explícita la reconstrucción de los parciales superpuestos (el bloque más delicado de la técnica propuesta) convendría explicar previamente los dos bloques adicionales necesarios: la detección de onsets en piezas musicales y la estimación de frecuencias fundamentales en señales multipitch.

4.5. Detección y localización de onsets

El ritmo de una pieza musical queda definido principalmente por los instantes de tiempo en que se producen los diferentes eventos sonoros (entrada de notas musicales, lírica,... etc.). La detección automática de los instantes de inicio de tales eventos u *onsets*, es muy importante también en compresión, codificación o segmentación de piezas y puede facilitar el camino para que ciertos efectos estándar en edición de audio (por ejemplo *pitch shifting* y *time stretching*) se adapten mejor a la señal.

La envolvente de una onda sonora suele simplificarse en cuatro pasos (envolvente ADSR): Ataque, Decaimiento, Sostenimiento y Relajación (en inglés *Attack, Decay, Sustain* y *Release*). El ataque es el tiempo que tarda el sonido en alcanzar su máxima amplitud. El decaimiento es el tiempo que tarda en alcanzar el nivel de sostenimiento, en el que se mantendrá mientras se mantenga presionada la tecla o se frote una cuerda o se sople un instrumento de viento. La relajación es el tiempo que tarda en desaparecer el sonido una vez que cesa la excitación. En la Figura 4.4 aparece la forma de onda de una nota de clarinete (en azul claro). Señaladas sobre la envolvente aproximada (en rojo) aparecen delimitadas de forma orientativa estas cuatro partes de la onda.

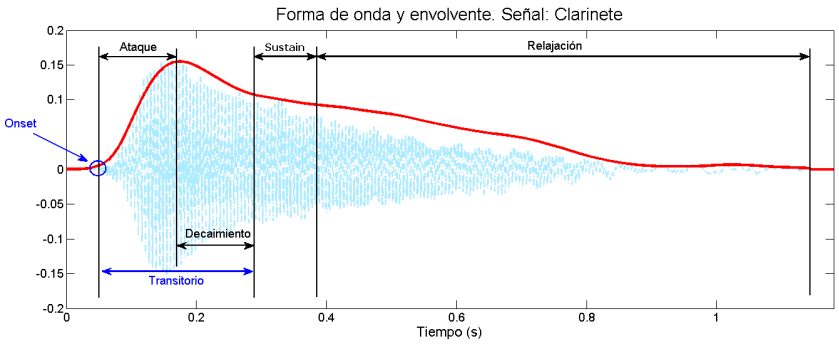


Figura 4.4: *Partes de la envolvente ADSR para una nota de clarinete.*

El transitorio es el instante de tiempo en el que se produce una mayor variabilidad en $x(t)$. Comprende básicamente el ataque y el decaimiento. El onset de una nota (en la Figura 4.4, señalado con un círculo azul) es idealmente el instante de tiempo en el que arranca

su transitorio. Diferentes instrumentos poseen distintos mecanismos de excitación y por lo tanto los onsets pueden poseer características muy heterogéneas. Una pieza musical genérica está compuesta por varias voces que pueden o no actuar en consonancia.

En esta Sección se va a presentar una técnica de localización de onsets basada en el algoritmo CWAS [23] cuyos resultados son prometedores. En primer lugar se repasarán muy brevemente distintos métodos existentes de detección de onsets. A continuación se detallará el método desarrollado, el cual ha sido puesto a prueba con cuatro señales musicales: una batería y un piano como piezas de gran contenido percusivo y transitorios muy bruscos, y un violín y una guitarra como ejemplos de una señales más armónicas. Los resultados así como la valoración de los mismos se presentan al final de la Sección 4.5.3.

4.5.1. Técnicas de localización de Onsets

Los métodos de detección automática de eventos musicales se basan en dos partes clave: la obtención a partir de la señal de audio $x(t)$ de una *función de detección* apropiada, y la selección a partir de ella de los picos asociados a un onset, mediante los llamados algoritmos de *peak picking*.

El primer candidato a función de detección que puede plantearse es la propia $x(t)$; pero su alta variabilidad y su casuística la hacen en general inapropiada para el objetivo que se persigue. Un algoritmo de búsqueda general debe basarse en una parte de la señal que guarde información de los eventos, pero poco más. De este modo, diferentes métodos de búsqueda comienzan por diferentes formas de obtener tal función de detección (preprocesado de la señal). En general, la búsqueda de onsets pasa tarde o temprano por un submuestreado de $x(t)$, reduciendo su variabilidad pero manteniendo la información de los onsets de la señal original. La función de detección puede ser generada de formas diferentes, basándose en técnicas de procesado temporal o frecuencial de la señal. Un resumen de las diferentes técnicas puede encontrarse en [16].

La envolvente temporal de $x(t)$ suele aumentar de forma más o menos brusca con la entrada de nuevas notas. Su energía (básicamente la amplitud cuadrática en cada instante de tiempo) se comporta de forma similar. Ambos métodos funcionan aceptablemente cuando se trata de detectar transitorios fuertemente percusivos sobre un fondo adecuadamente libre de ruidos, pero presentan cierta tendencia a señalar onsets espurios. Otros métodos emplean la psicoacústica según la cual el oído humano funciona de forma logarítmica [89]. Un cambio brusco en la derivada primera del logaritmo de la energía $d(logE)/dt$, simula la percepción auditiva de la entrada de una nota.

En cuanto al tratamiento espectral de la información, tiene tendencia a reducir la necesidad de submuestrear $x(t)$. Existen varios métodos de detección basados en el contenido frecuencial de la señal y la gran mayoría de ellos emplea la STFT para obtener la infor-

mación tiempo-frecuencia que la caracteriza. Algunos métodos se aprovechan de que los transitorios quedan especialmente marcados en las zonas de alta frecuencia, mientras que los cambios en la energía suelen concentrarse en zonas de baja frecuencia. Otros emplean la evolución espectral de la señal para detectar cambios y con ellos los correspondientes onsets. También se pueden detectar eventos atendiendo a la información implícita en la fase: en un transitorio, la derivada segunda de la fase sufre un cambio brusco que puede ser detectado, considerando como promedio estadístico el estacionario de cada nota ejecutada.

Como se ha adelantado, encontrar una función de detección adecuada, aunque crítico, es sólo un paso en el proceso de localización. Una vez obtenida, se han de llevar a cabo el proceso de peak picking y la posterior identificación de los onsets. Existen multitud de algoritmos de peak picking, y múltiples formas de identificar los onsets adecuadamente. Cabe destacar el uso de umbrales adaptativos [87] por ser similar la solución aquí empleada, como se verá más adelante.

4.5.2. El algoritmo CWAS como localizador de Onsets

A continuación se describirá brevemente el algoritmo de detección de onsets desarrollado, cuyos bloques principales pueden apreciarse en la Figura 4.5.

Como se puede ver, se parte del análisis de la señal de audio que proporciona el algoritmo CWAS. A continuación se procede a la búsqueda de la función de detección, que en este caso va a partir de la información contenida la zona de alta frecuencia (bandas inferiores) del análisis. El algoritmo inicial de detección marca las posiciones de los onsets de la función de detección. La tendencia es a marcar bastantes más onsets de los realmente existentes, de modo se hace necesario llevar a cabo una selección posterior, utilizando para ello una adecuada umbralización. El resultado es la asignación de los onsets correctos, cuya localización exacta en el tiempo habrá que afinar antes de ofrecer los datos como salida.

4.5.2.1. Técnica preliminar de detección

Una vez obtenidos los coeficientes wavelet de la señal $x(t)$, se procede a la obtención de la función de detección. En este caso, será:

$$f_{D_x}(t) = \sum_{j=1}^{N} \|W_x(t_i, k_j)\|^2 \tag{4.9}$$

Por comodidad, se escribe $f_{D_x}(t)$ si bien es evidente que el tiempo está discretizado. Concretando, la función de detección inicial escogida presenta la información contenida en las $N = 10$ bandas inferiores del espectro. Así $f_{D_x}(t)$ conserva el contenido transitorio de las señales utilizadas (aunque tales transitorios no siempre se corresponden con el ataque

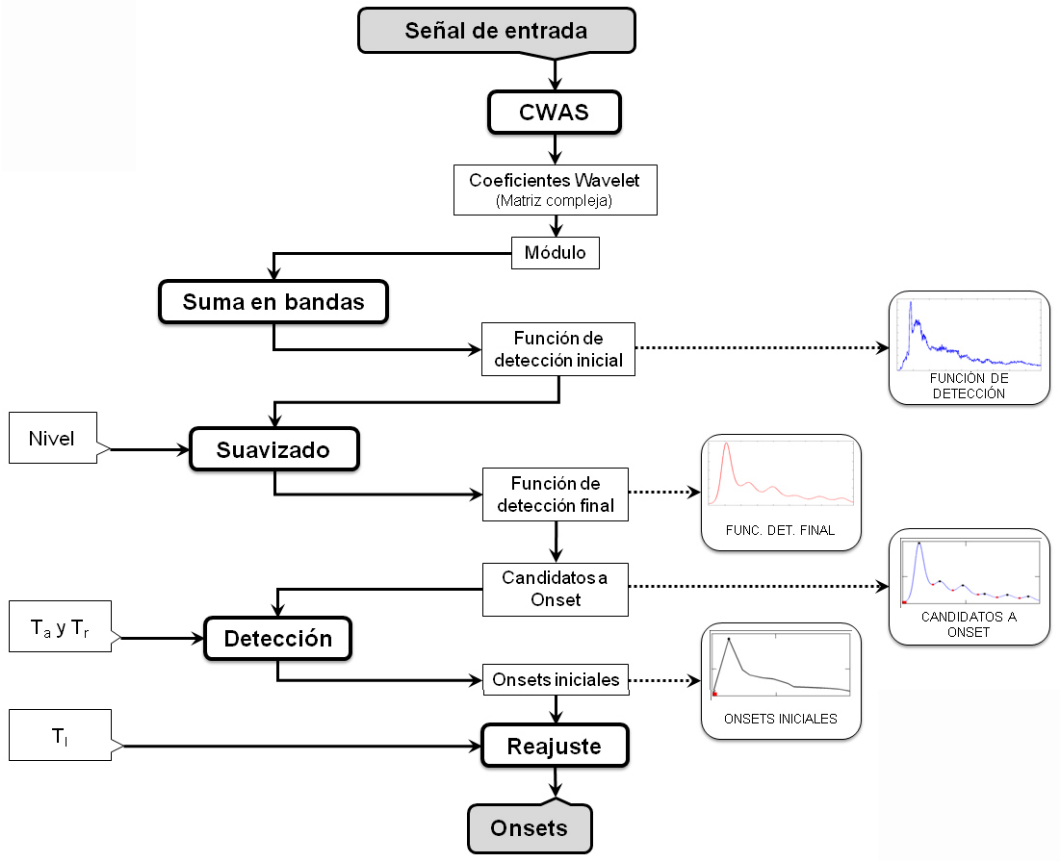


Figura 4.5: *Diagrama de bloques del algoritmo de onsets basado en CWAS.*

de notas o con golpes de percusión, como se verá en los resultados, más adelante). En efecto, los transitorios quedan mejor marcados en las frecuencias superiores, aunque en las notas armónicas de menor frecuencia puedan dejar una huella más débil. La inclusión de un número de bandas superior $N > 10$ puede provocar la aparición de onsets espurios en alguna de las señales cuyos resultados se muestran en la Sección 4.5.3, sin embargo esta técnica presenta una efectividad y una precisión más que suficientes para el uso que se va a hacer de ella, que consiste en una estimación de los tiempos de inicio y fin de ejecución de una nota musical.

4.5.2.1.1. Función de detección

La función $f_{D_x}(t)$ todavía presenta un exceso de variabilidad. Es necesario suavizarla, antes de proceder a la búsqueda sistemática de picos. Este suavizado se puede llevar a cabo

de múltiples formas: mediante la unión de los máximos de la función f_{D_x}, a través de un filtrado pasobajo suficiente de la misma, o, como se ha decidido en este caso, calculando el equivalente a la envolvente espectral de la señal en el dominio del tiempo mediante una envolvente *cepstral*[1] [33]:

$$f_{FD_x}(t) = IFFT\Big\{w_{LP}(\omega, N_1) \cdot FFT\big[log(f_{D_x})\big]\Big\} \tag{4.10}$$

En esta expresión, $w_{LP}(\omega, N_1)$ es un filtro pasobajo tal que:

$$w_{LP}(\omega, N_1) = \begin{cases} 1 & \text{si } \omega = 1, N_1 \\ 2 & \text{si } 1 \leq \omega < N_1 \\ 0 & \text{si } N_1 < \omega \leq M-1 \end{cases} \tag{4.11}$$

donde M es el número de muestras de la señal.

El resultado de la Ecuación (4.10) es la señal final de detección, $f_{DF_x}(t)$. En ella se mantienen las cualidades transitorias más importantes de la señal, tanto más deslocalizadas temporalmente cuanto mayor sea el nivel de filtrado, es decir, menor N_1 en la Ecuación (4.10).

Aunque presenta ciertas características ventajosas como su mayor adaptabilidad en algoritmos más elaborados, se ha comprobado que la envolvente cepstral no supone, al menos para las señales analizadas, resultados muy diferentes a los que se obtendrían con, por ejemplo, un simple filtrado pasobajo de la señal. Por otro lado, la unión simple de los máximos en $f_{D_x}(t)$ suele presentar todavía un exceso de variabilidad superflua que obliga a una repetición escalonada o a un posterior filtrado pasobajo.

4.5.2.1.2. Detección de picos

Una vez obtenida la función de detección final, $f_{DF_x}(t)$, se trata de localizar los picos dentro de la misma. Evidentemente, es necesaria la presencia de un pico tras cada onset (que marca el máximo del ataque de cada nota presente en $x(t)$, algo no siempre distinguible en la forma de onda de la señal), si bien no todos los picos de $f_{DF_x}(t)$ indican la entrada de un evento. En la Figura 4.6 se ha representado una parte de $f_{DF_x}(t)$ para el caso de una señal de piano. En la gráfica se puede observar todo un conjunto de picos asociados a $f_{DF_x}(t)$, cada uno de ellos asociado a un valle. De todo el conjunto inicial, tan sólo la primera pareja valle-pico (destacada con un tamaño mayor) se corresponde con un verdadero onset. Para distinguir los valles que representan auténticos onsets de los generados por ejemplo por

[1] En el trabajo original de Bogert, Healy y Tukey mencionado en la bibliografía, el *cepstrum* de una señal es el resultado de calcular la FT del logaritmo del espectro de la misma. Es decir, se trata tal función como si fuese una señal de audio por derecho propio.

efectos de *bending* o *vibratos*, o por la simple naturaleza del instrumento musical, se han introducido sendos parámetros de control, T_a y T_r, en un algoritmo adaptativo.

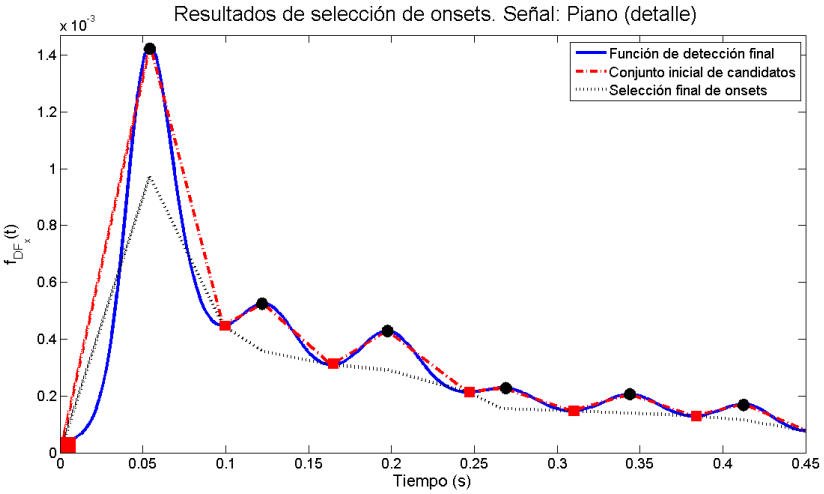


Figura 4.6: *Conjunto de valles y picos asociados a la función $f_{DF_x}(t)$ para una señal de piano.*

Sean P y V los conjuntos de picos y valles de $f_{DF_x}(t)$ respectivamente. En la Figura 4.6, P sería el conjunto de puntos negros y V el de cuadros rojos. Sólo los valles asociados a picos suficientemente abruptos son indicativos de onset. Utilizando los umbrales T_r y T_a (uno relativo y otro absoluto), es posible extraer V_1, el subconjunto de onsets adecuados de V.

Forzando:

$$T_r \cdot p_{j,j-1} > v_j \quad \forall v_j \epsilon V \tag{4.12}$$

se seleccionarán únicamente aquellos valles situados entre picos que sean lo suficientemente elevados respecto su valor. En esta ecuación, p_j y p_{j+1} son las alturas de dos picos consecutivos, mientras que el valor de v_j es el del valle situado entre ellos. De este modo se procede a un agrupamiento de valles asociados a un mismo onset.

A continuación, de los valles que cumplan con la Ecuación (4.12) y que pertenezcan por lo tanto al subconjunto V_1, se seleccionarán aquellos asociados a picos que posean una energía suficiente con respecto al máximo absoluto de la señal, es decir:

$$p_k > T_a \cdot max\left\{p_j \epsilon P\right\} \quad \forall v_k \epsilon V_1 \tag{4.13}$$

La posición k (marca temporal) de los valles de $f_{DF_x}(t)$ que cumplan con las Ecuaciones

(4.12) y (4.13) serán considerados como onsets válidos.

4.5.2.1.3. Relocalización

Para fijar la posición final de cada onset sobre la forma de onda, se lleva a cabo un pequeño proceso de reajuste, con el objetivo de corregir la deslocalización provocada por el proceso de filtrado pasobajo de $f_{D_x}(t)$ para obtener $f_{DF_x}(t)$. El resultado aparece reflejado en la Figura 4.7. En esta figura se han presentado los resultados de reajuste de un onset de la señal del piano (gráfica superior). En la gráfica inferior, una ampliación de la misma.

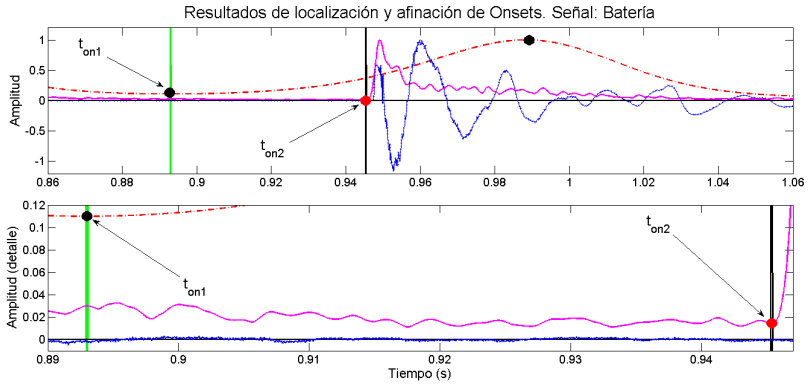


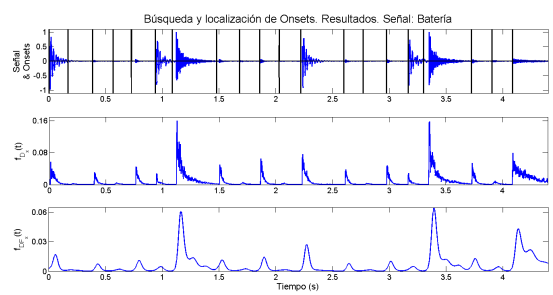
Figura 4.7: *Localización fina de onsets. En azul, la forma de onda original. En tono rosado, la función de detección inicial. En rojo, la función de detección final. La posición inicial del onset está marcada como* t_{on1} *(trazo vertical verde). El proceso de afinamiento la traslada hasta* t_{on2} *(trazo vertical negro).*

En la gráfica superior de la Figura 4.7 se puede apreciar cómo entre la posición inicial del onset (t_{on1}, en un mínimo de $f_{DF_x}(t)$, marcado con trazo discontinuo) y su máximo siguiente (ambos puntos resaltados en negro en la gráfica), se debe localizar el mínimo de $f_{D_x}(t)$ (en rosa) más cercano a este máximo, dentro de unos límites de tolerancia marcados por un umbral T_l (último umbral algorítmico, en la Figura 4.5). El resultado de este ajuste es la posición optimizada del onset, t_{on2} (punto rojo en la figura).

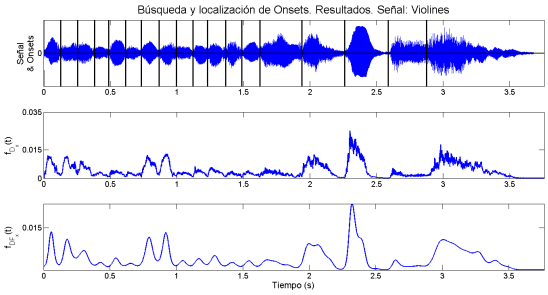
En la ampliación inferior de la Figura 4.7 se pueden observar en detalle las posiciones inicial y final del onset, sobre mínimos respectivos de la función de detección suavizada $f_{DF_x}(t)$ (trazo discontinuo rojo) y la función de detección inicial $f_{D_x}(t)$ (trazo continuo rosa). Pese a todo, la localización de los onsets todavía no es exacta, pues tiende a presentar un pequeño desplazamiento temporal a la izquierda de su posición ideal (es decir, los marcadores finales están situados en realidad un poco antes de sus posiciones óptimas).

4.5.3. Resultados y valoración

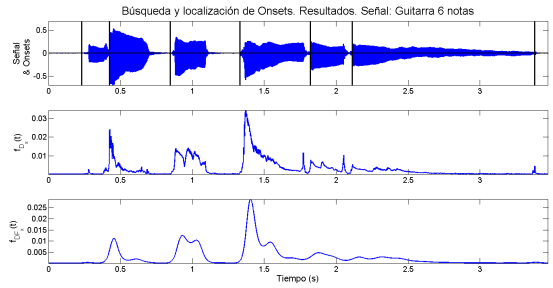
Se ha puesto a prueba éste algoritmo con un conjunto de cuatro señales de audio, como se ha dicho antes, pasajes de violín, guitarra, piano y batería, intentando cubrir un abanico de posibilidades ilustrativo. Los resultados obtenidos se muestran en las Figuras 4.8(a) a 4.8(d). La elección de todos los parámetros de control y la función de detección se ha llevado a cabo con el objetivo de generar un algoritmo lo más adaptable posible.



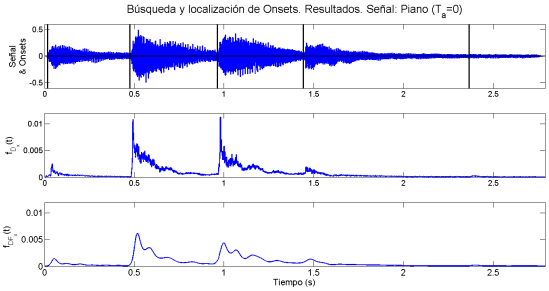
(a) *Onsets: batería.*



(b) *Onsets: violines.*



(c) *Onsets: guitarra.*



(d) *Onsets: piano.*

Figura 4.8: *Resultados finales de búsqueda y localización de onsets: (a) Señal de batería. (b) Señal de violines. (c) Señal de guitarra. (d) Señal de piano. En cada figura: Gráfica superior, forma de onda (azul) y onsets (negro). Centro, función de detección* $f_{D_x}(t)$. *Gráfica inferior, función de detección final* $f_{DF_x}(t)$.

En la parte superior de cada sub-figura, (a), (b), (c) y (d), se muestra la forma de onda correspondiente a cada señal, junto con los onsets detectados (en negro). A continuación, la función de detección $f_{D_x}(t)$, obtenida a través de la Ecuación (4.9). Aplicando sobre esta la Ecuación (4.10), se obtiene $f_{DF_x}(t)$, que aparece representada en la gráfica inferior.

En el proceso de obtención de $f_{DF_x}(t)$, a través de la Ecuación (4.10), se ha tomado como nivel de filtrado $N_1 = 80$. Por último, los umbrales absoluto y relativo que controlan el algoritmo de selección de onsets se han tomado también idénticos. En este caso $T_a = 0$

y $T_r \approx -3dB$. En cuanto a la fiabilidad de resultados en la selección de onsets, es de un 100 % para las señales estudiadas. Es decir, se marcan todos y cada uno de los onsets audibles en cada caso, si bien en las formas de onda de las Figuras 4.8(c) y 4.8(d) parecen presentar sendos espurios. En el caso de la guitarra, se trata de un onset producido por la púa al rozar la cuerda. En el caso del piano, es el ruido de un golpeo final del martillo al retirarse. En la representación de $f_{D_x}(t)$ de ambas señales (gráficas centrales), estos golpes quedan bastante marcados (especialmente el de la guitarra). Aunque ambos sonidos resultan realmente audibles, considerar estos eventos como onsets es al menos cuestionable, si bien sus correspondientes marcadores pueden ser eliminados eligiendo un umbral absoluto no nulo (concretamente, $T_a = 4\,\%$).

4.6. Estimación de frecuencias fundamentales

En los instrumentos musicales, incluso al ejecutarse una nota afinada, la excitación no produce un tono puro sino todo un conjunto de frecuencias con sus correspondientes amplitudes. El conjunto de envolventes de las componentes es lo que permite al cerebro distinguir entre diferentes instrumentos (timbre), mientras que la nota que se está ejecutando queda caracterizada por la *frecuencia fundamental* o *pitch* del sonido. En la distribución armónica de un sonido, la frecuencia fundamental es la frecuencia base a partir de la cual se generan (idealmente) todos los demás armónicos presentes. En general no tiene por qué corresponderse con el parcial de mayor amplitud, ya que alguno de sus armónicos puede poseer más energía.

Evidentemente, las frecuencias fundamentales presentes en una mezcla son un parámetro mucho más interesante de cara a distinguir cuántas fuentes coexisten y qué información le corresponde a cada una. La técnica de separación propuesta en la sección anterior emplea la detección de fundamentales para estos objetivos, pero el método de estimación empleado es, en el mejor de los casos, demasiado grosero y puede fallar en mezclas más complicadas, por ejemplo de tres fuentes, con instrumentos más ruidosos o que presenten información subarmónica. Por lo tanto, se hace necesario el desarrollo de un estimador de frecuencias fundamentales en señales multipitch más fiable de cara a obtener los mejores resultados.

A continuación, se expondrán dos algoritmos de detección de frecuencias fundamentales. Se trata de una propuesta inicial basada exclusivamente en criterios energéticos y armónicos, y de una propuesta más completa, que incluye posibilidades no tenidas en cuenta inicialmente, como la posibilidad de relaciones armónicas perfectas entre fundamentales o la posibilidad de detectar fundamentales suprimidas. Ambas técnicas han sido desarrolladas para separar instrumentos musicales monofónicos. Caso de analizarse instrumentos polifónicos (por ejemplo un piano ejecutando un acorde), cada nota presente de ese instrumento será tratada como una fuente independiente.

4.6.1. Técnica inicial

El algoritmo CWAS, con su caracterización coherente de amplitudes y fases (frecuencias), resulta ideal para localizar e identificar la frecuencia fundamental de la señal analizada. En la Figura 4.9 se presenta el diagrama de bloques propuesto para encontrar el pitch en una señal polifónica, partiendo de la información proporcionada por el algoritmo CWAS.

La técnica de búsqueda puede explicarse como sigue: en un contexto de cálculo frame-to-frame, se parte de la información en máscaras bidimensionales obtenida en el algoritmo CWAS básico, a partir de las cuales se generan los diferentes parciales complejos detectados por el filtrado pasobanda. Tales parciales son analizados en términos de módulo y fase (de la cual a su vez se extrae la frecuencia instantánea). Los parciales son ordenados según criterios energéticos, y sólo los más importantes son tenidos en cuenta a la hora de buscar la frecuencia fundamental. Es decir, aquellos parciales cuya energía, ya sea por extensión temporal del parcial o por amplitud insuficientes, no alcance un cierto límite porcentual respecto a la cantidad de energía total contenida en el frame, no serán tenidos en cuenta en el cálculo. Los parciales energéticamente importantes se ordenan de mayor a menor. Para cada uno de ellos se calcula su frecuencia promedio $\bar{f}_n$, utilizando la expresión:

$$\bar{f}_i = \bar{f_{ins,i}} = \frac{1}{\|t_i\|} \sum_{t_m} f_{ins,i}(t_i), \qquad \forall t_m | A_i(t_m) > \theta_1 \tag{4.14}$$

donde θ_1 es un umbral energético utilizado para evitar que los puntos donde la amplitud del parcial es demasiado baja entren en el cálculo de $\bar{f}_i$ (ya que la frecuencia instantánea no está bien definida en tales puntos, y presenta grandes variaciones respecto a los demás). Se designa por $\|t_i\|$ al cardinal de este conjunto de puntos, es decir, el número de muestras del parcial que satisfacen el citado criterio umbral.

Comenzando por el parcial de mayor energía, se realiza un *matching de armónicos*. Para ello, se calcula la relación entre la frecuencia promedio del parcial y las frecuencias promedio de cada uno de los demás parciales detectados. Un parcial j se considerará armónico puro de otro, i, cuando el ratio entre las frecuencias instantáneas de ambos sea un número entero positivo (con una tolerancia controlada por un nuevo umbral[2]):

$$\frac{\bar{f}_j}{\bar{f}_i} \approx n \qquad n \in \mathbb{N} \tag{4.15}$$

Es decir, no serán tenidos en cuenta los parciales proporcionales a otros con proporcionalidad no entera. Esto no supone renunciar a ninguna multiplicidad, ya que el barrido se efectúa para todos los parciales energéticamente importantes de la señal, y la frecuencia fundamental está ligada al parcial que ofrece un matching de armónicos más elevado, aquel

[2] *Véase la Sección 4.6.2 para un análisis más detallado.*

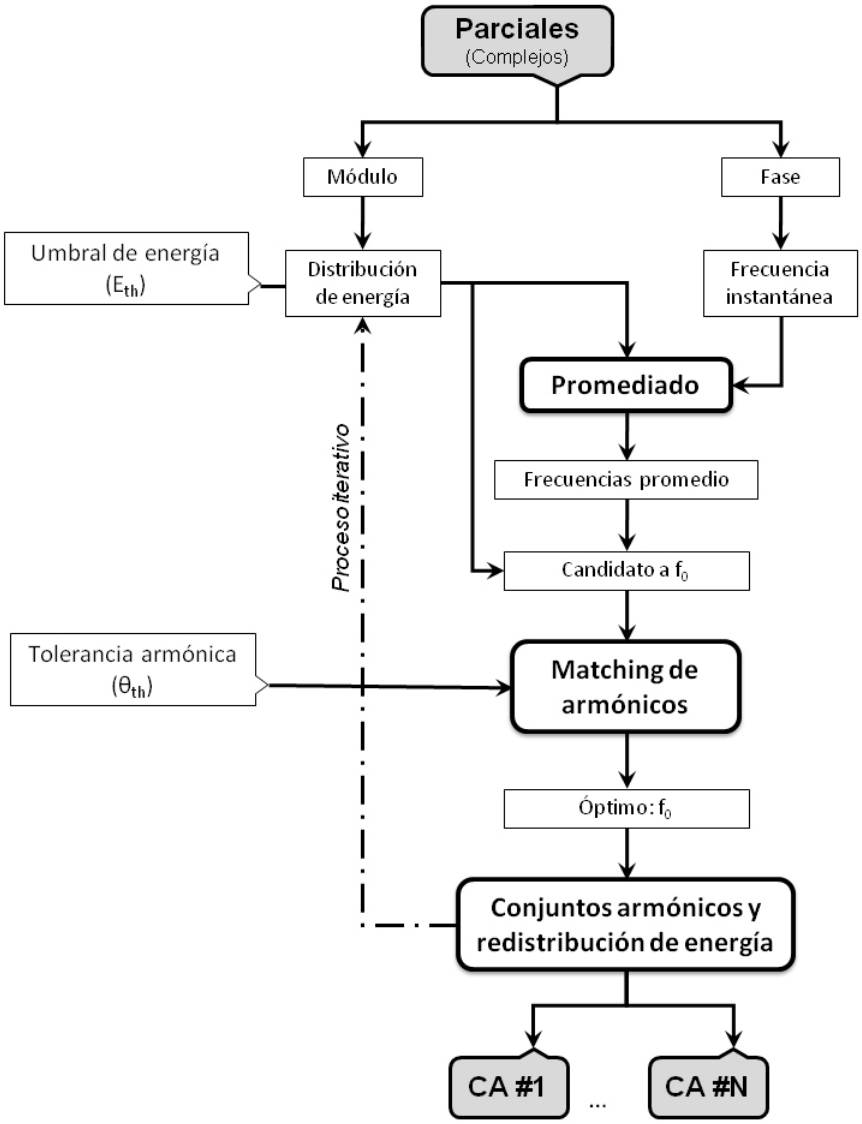


Figura 4.9: *Primer algoritmo de detección de frecuencia fundamental en señales multi-pitch. Las salidas CA#1 a CA#N son los conjuntos armónicos (peines frecuenciales) relacionados con cada una de las N fuentes detectadas.*

que presenta un mayor número de armónicos definidos a través de la Ecuación (4.15).

Este proceso es iterativo, por lo que se hace posible detectar varios tonos sonando simultáneamente. Una vez obtenida una frecuencia principal, se buscará otra entre los parciales restantes (no armónicos respecto a la f_0 detectada) una segunda posible frecuencia

principal, repitiendo a continuación para este nuevo candidato el proceso de matching de armónicos. En este algoritmo no se han tenido en cuenta dos casos de fundamental importantes: fundamentales caracterizadas por una relación armónica pura (por ejemplo, f_{01} =220Hz y f_{02} =440Hz) y fundamentales suprimidas (cuando aparecen los armónicos correspondientes a una cierta fundamental, pero el parcial fundamental en sí no existe).

4.6.2. Algoritmo propuesto

El algoritmo final de estimación de frecuencias fundamentales [18] ha sido desarrollado para cubrir estas carencias básicas. Se trata de un algoritmo que emplea la técnica denominada suavizado espectral [90, 91], la cual ha debido ser especialmente adaptada para analizar de forma correcta la información de salida del algoritmo CWAS. En la Figura 4.10 se muestra el diagrama de bloques del programa desarrollado.

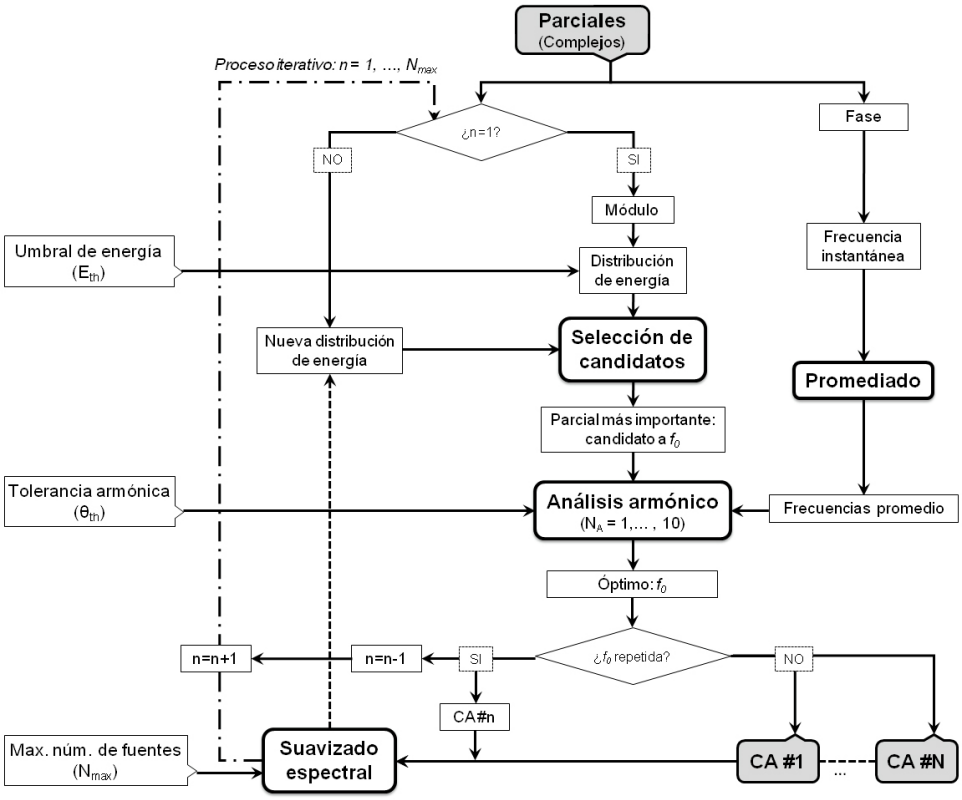


Figura 4.10: *Diagrama de bloques del algoritmo final de estimación de frecuencia fundamental en señales multipitch. Las salidas $CA\#1$ a $CA\#N$ son los conjuntos armónicos (peines frecuenciales) relacionados con cada una de las N fuentes detectadas.*

La señal de entrada (mezclada) se analiza utilizando el algoritmo CWAS, que ofrece como resultados las n funciones complejas que definen la evolución temporal de cada parcial

detectado. Utilizando las ecuaciones adecuadas, se obtienen las frecuencias instantáneas para cada parcial $f_j(t)$ y sus respectivos valores medios, $\overline{f_j}$, $\forall j = 1, \ldots, n$), así como la distribución de energía de la señal (es decir, la energía de cada parcial). Esta información es equivalente a la de la señal de escalograma, pero evaluado únicamente para el conjunto de parciales detectados. Los diferentes parciales se subdividen en dos categorías, en función de un umbral de energía, en este caso $E_{th} = 1\,\%$. Partiendo de la distribución de energías, el parcial más enérgico es seleccionado para realizar a partir de éste el análisis armónico.

Suponiendo que la frecuencia media del parcial más energético es $\overline{f_j}$, se asume que este parcial es a su vez un armónico de cierto orden de una frecuencia fundamental f_{0k} desconocida, es decir:

$$f_{0k} = \frac{\overline{f_j}}{k}, \qquad \forall k = 1, 2, \ldots, N_A \tag{4.16}$$

El algoritmo se limita a estudiar los armónicos en potencia hasta un nivel $N_A = 10$. Dicho de otro modo, el parcial seleccionado será como mucho el 10° armónico de su correspondiente frecuencia fundamental. Del conjunto de candidatos a frecuencia fundamental así obtenidos, se calcula a su vez el peine de frecuencias armónicas correspondiente para cada una de ellas:

$$f_{k,m} = m f_{0k}, \qquad \forall m = 1, 2, \ldots, N_k \tag{4.17}$$

donde N_k es el natural más grande que satisface $N_k f_{0k} \leq f_s/2$, siendo f_s la frecuencia de muestreo.

En el siguiente paso, para cada $f_{k,m}$, se busca su correspondiente parcial. Un parcial de frecuencia promedio $\overline{f_i}$ es el $m - simo$ armónico de cierta frecuencia fundamental f_{0k} si:

$$\left| \frac{\overline{f_i}}{f_{0k}} - \frac{f_{k,m}}{f_{0k}} \right| \leq \theta_a \tag{4.18}$$

donde en este caso θ_a es el umbral de *inarmonicidad*. Tomando θ_a=0.03, incluso los parciales de un instrumento musical inarmónico como el piano son correctamente seleccionados.

La decisión acerca de cual es la frecuencia fundamental ligada con el parcial j bajo estudio se toma a través de una función de peso diseñada específicamente para esta aplicación, w_k. Esta función de peso está basada en dos criterios: la energía total de cada peine armónico involucrado y la presencia o ausencia de los diferentes armónicos.

En concreto:

$$w_k = \frac{n_{ip,k}^2}{n_{a,k}} \sum_{i=1}^{n_{a,k}} E_{i,k} \tag{4.19}$$

donde $n_{a,k}$ es el número total de armónicos asociados con la fundamental f_{0k} y $n_{ip,k}$ es el número de parciales cuya energía supera el límite E_{th}. Por su parte, $E_{i,k}$ es la energía del $i - simo$ parcial de f_{0k}.

La fundamental seleccionada es aquella cuyo peso w_k es máximo. El algoritmo almacena el conjunto de parciales armónicos o *patrón espectral* de la fuente, $\mathbf{P}_k = \{P_{1,k}, P_{2,k}, \dots P_{n_a,k}\}$, que por supuesto incluye la frecuencia fundamental seleccionada, y a continuación se aplica el suavizado espectral [91] a la distribución de energía de la fuente, $\mathbf{E}_k = \{E_{1,k}, E_{2,k}, \dots E_{n_a,k}\}$:

$$\widetilde{\mathbf{E}}_k = G_w \otimes \mathbf{E}_k \tag{4.20}$$

donde $G_w = \{0.212, 0.576, 0.212\}$ es una ventana Gaussiana normalizada truncada a tres componentes y $\otimes$ es el operador producto de convolución. La energía suavizada para cada parcial armónico se calcula como:

$$E'_{i,k} = \begin{cases} E_{i,k} - \widetilde{E_{i,k}} & \text{si } E_{i,k} - \widetilde{E_{i,k}} > 0 \\ 0 & \text{si } E_{i,k} - \widetilde{E_{i,k}} \leq 0 \end{cases} \tag{4.21}$$

Sustituyendo los nuevos valores de energía de los parciales del conjunto armónico en la distribución original de energía, se puede buscar un nuevo parcial preponderante y repetir el procedimiento de forma iterativa. Cuando la energía restante de la distribución queda por debajo de un límite o se alcanza el número máximo de fuentes (en este caso limitado a 5), el proceso se detiene. Con esta técnica es posible obtener una estimación de frecuencias fundamentales incluso en casos complicados, por ejemplo cuando una fundamental se superpone con un armónico de otra o en el caso de fundamentales suprimidas. Las fundamentales superpuestas no se detectan empleando esta técnica.

4.6.3. Resultados y valoración

Este algoritmo ha sido testado sobre un conjunto de más de 200 señales de instrumentos musicales reales, la mayoría de los cuales han sido obtenidos una vez más de la base de datos de la Universidad de Iowa [63]. Los ensayos comprenden 4 categorías diferentes: instrumentos musicales aislados, y mezclas sintéticas de dos y tres fuentes, y de un instrumento armónico mas un instrumento inarmónico (piano). Los resultados se resumen en la Tabla 4.1.

Resultados de estimación de frecuencias fundamentales			
	Señales analizadas (#)	**Aciertos (#)**	**Error (%)**
1 instr.	106	106	0
2 instr.	75	74	1.34
1 inst. A+1 inst. I	4	4	0
3 instr.	50	49	2
TOTAL	235	233	0.85

Tabla 4.1: *Resultados de precisión en el algoritmo de estimación de frecuencias fundamentales.*

En general, los errores en la estimación pueden deberse a detecciones fallidas (cuando una fundamental presente no es detectada), fundamentales erróneas (cuando aparece una fundamental espuria) o estimaciones erróneas (cuando una fundamental aparece desplazada respecto de su valor teórico). En el caso de nuestro algoritmo, los errores que aparecen en la Tabla 4.1 son de las dos primeras categorías, y pueden eliminarse empleando un umbral de energía diferente para las señales involucradas.

4.7. Algoritmo de separación de notas musicales

En la última técnica de separación que se detallará a continuación, se emplea la gran capacidad del algoritmo CWAS para obtener las amplitudes y fases instantáneas de cada parcial detectado para relajar ligeramente la aproximación de CAM. Los armónicos no superpuestos se obtienen mediante el adecuado enmascaramiento del espectrograma wavelet. En cuanto a los parciales superpuestos, las amplitudes instantáneas de cada fuente mezclada se reconstruyen totalmente partiendo de las de parciales aislados mediante una proporcionalidad directa siguiendo criterios energéticos mediante una aproximación por mínimos cuadrados, como se explicitará más adelante. De esta forma, es posible relajar la alta restricción en la obtención de las fases exactas de cada contribución, y éstas pueden ser asimismo reconstruidas partiendo de la información de fase de parciales no superpuestos.

El algoritmo de esta aproximación final a la separación se ha representado de forma esquemática en la Figura 4.11.

Se parte de la información arrojada por el algoritmo de búsqueda de fundamentales presentado en la Sección 4.6, consistente en los conjuntos de parciales armónicos asociados a cada una de las N fuentes de la mezcla. El proceso es *off-line*, posterior al análisis frame-to-frame y al tracking de parciales, y se lleva a cabo empleando la información global de la señal.

Con esta información se pueden distinguir dos tipos de parciales:

1. Parciales aislados, es decir, parciales armónicos de una única fuente.

2. Parciales superpuestos, aquellos que son armónicos de más de una fuente.

Por lo tanto, se ignorará el conjunto de parciales no armónicos del algoritmo CWAS (lo cual afectará evidentemente a la calidad numérica y sonora de ciertos instrumentos). Los parciales aislados serán utilizados para estimar las contribuciones presentes en los parciales compartidos utilizando envolventes, fases y tiempos de onset y offset. Cada fuente separada se sintetiza finalmente sumando el conjunto de sus parciales asociados (aislados y separados), tras lo cual se llevará a cabo la medida de calidad de separación (Sección 4.7.6).

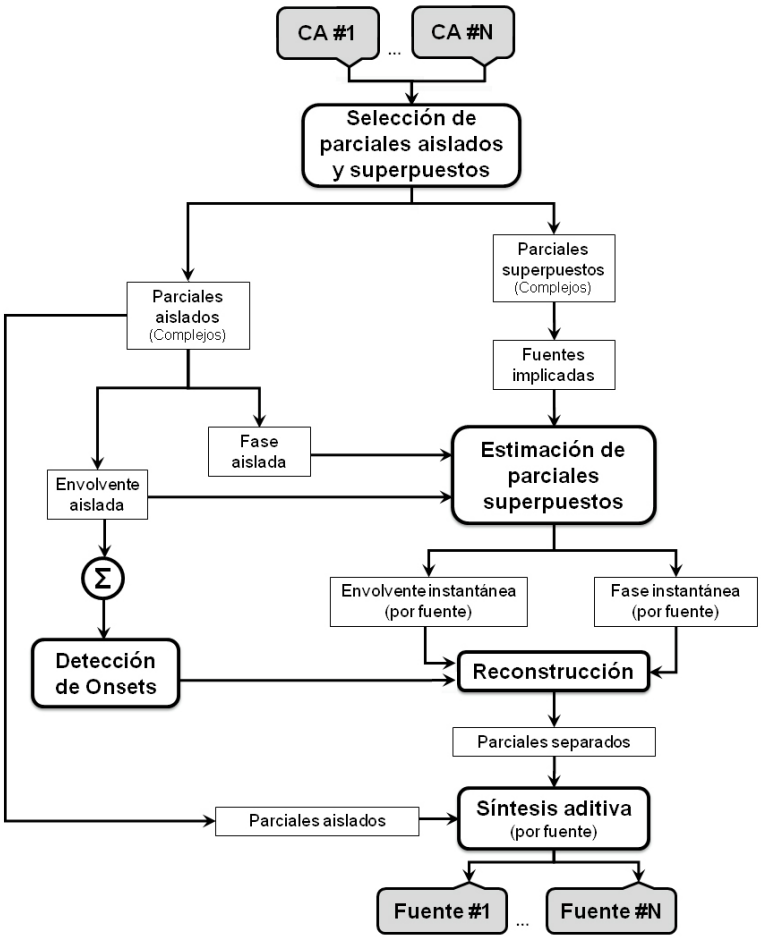


Figura 4.11: *Diagrama de bloques del algoritmo de separación monaural de parciales superpuestos.*

4.7.1. El límite inarmónico

La inarmonicidad es un fenómeno que ocurre principalmente en instrumentos de cuerda debido a la rigidez de la cuerda y a sus terminaciones no rígidas. Como resultado, cada parcial presenta una frecuencia mayor que su correspondiente valor armónico. Por ejemplo, la ecuación de la inarmonicidad para un piano puede ser escrita [128] como:

$$f_n = n f_1 \frac{\sqrt{1+\beta n^2}}{\sqrt{1+\beta}} \tag{4.22}$$

donde f_1 es la frecuencia del primer parcial presente en el espectro (escala temperada), n es el número armónico y β es el parámetro de inarmonicidad. En la Ecuación (4.22), β se

asume constante, aunque se puede modelar de forma más precisa por medio de un polinomio que puede llegar a tener grado 7 [127]. Esto significa que se obtienen diferentes valores para este parámetro dependiendo de los parciales que se empleen para calcularlo. Los parciales situados en las octavas 6-7 arrojan resultados óptimos. Por medio de dos parciales de orden m (inferior) y n (superior), se tiene:

$$\beta = \frac{\delta - \varepsilon}{\varepsilon n^2 - \delta m^2} \tag{4.23}$$

donde $\delta = (mf_n/nf_m)^2$ y ε es un error inducido debido a la estructura física del piano que no puede ser evaluado [128]. Si los parciales m y n son adecuadamente seleccionados, $\varepsilon \approx 1$.

Con el modelo inarmónico codificado en la Ecuación (4.23), es posible calcular el parámetro de inarmonicidad β para cada fuente detectada empleando (cuando sea posible) dos parciales aislados situados en las octavas apropiadas. A priori, esta técnica incluye los instrumentos inarmónicos (como el piano) en el modelo propuesto. Desafortunadamente, la obtención de β no mejora significativamente la calidad de la separación evaluada en las pruebas que se detallarán más adelante.

4.7.2. Supuestos

De cara a obtener las amplitudes y fases de un parcial superpuesto correspondientes a cada fuente, se asumen dos aproximaciones. La primera es una versión menos restrictiva del principio de Modulación de Amplitud Común (CAM, Sección 4.7), el cual afirma que las envolventes de las componentes espectrales de la misma fuente están correladas [110].

- Las amplitudes (envolventes) de dos armónicos P_1 y P_2, *con energía similar* $E_1 \approx E_2$, pertenecientes a la misma fuente presentan un alto coeficiente de correlación.

Cuanto más cierta sea esta aproximación, los resultados de separación serán tanto mejores. Como se emplea la información de la señal completa, los coeficientes de correlación entre el armónico más energético y los demás decrece a medida que las diferencias de amplitud/energía entre los parciales involucrados sean mayores [110]. El criterio de selección por similitud energética proporciona un medio para aumentar la calidad de los resultados numéricos y sonoros de separación. Si la energía del parcial mezclado es alta, se recurre a parciales aislados de gran peso energético, los cuales guardarán un alto coeficiente de correlación [110]. Si la energía de la mezcla es baja, del mismo modo se recurrirá a parciales de baja energía para separar (en la medida en que estos existan), los cuales tienden a dotar a la señal de un color similar a la original. Si no hay disponibles parciales aislados de baja energía, la tendencia es igualmente a no cometer errores numéricos demasiado elevados, si bien el parecido de las fuentes separadas respecto a los originales pasa a depender del número de señales mezcladas y del proceso de mezcla.

La segunda aproximación es:

- Las fases instantáneas de los parciales armónicos $p-simo$ y $q-simo$ pertenecientes a la misma fuente son aproximadamente proporcionales con una razón p/q, excepto un desfase inicial ϕ_0. En otras palabras:

$$\phi_2(t) \approx \frac{p}{q}\phi_1(t) + \Delta\phi_0 \tag{4.24}$$

donde $\Delta\phi_0 = 0$ significa que las fases iniciales de los parciales involucrados son iguales, es decir $\phi_{0p} = \phi_{0q}$.

Hemos encontrado que, en nuestro modelo de la señal de audio e incluso conociendo las envolventes de los parciales originales que se superponen, una diferencia de fase inicial $\Delta\phi_0 = 10^{-3}$ es suficiente para hacer imposible la adecuada reconstrucción de los parciales (para más información, consúltese el Anexo III.d). Esto es debido a que cada parcial presenta una fase inicial aleatoria (es decir, no hay relación aparente entre ϕ_{0p} y ϕ_{0q}). Sin embargo, dado que la frecuencia instantánea de los armónicos mezclados se puede recuperar con precisión independientemente del valor de la fase inicial, los parciales mezclados originales y los mezclados sintéticamente (a partir de la contribución de cada fuente separada) presentan sonidos similares (siempre que el primer supuesto sea cierto).

4.7.3. Proceso de reconstrucción y síntesis aditiva

Como se ha mencionado anteriormente, en la técnica propuesta se emplea la información de los parciales aislados para reconstruir la información de los superpuestos. La salida del algoritmo de estimación multipitch es el conjunto armónico correspondiente a cada fuente presente en la mezcla, Con esta información, una simple comparativa permite distinguir entre parciales aislados ($\mathbf{P}_k^{(iso)}$, pertenecientes a una única fuente) y parciales superpuestos o compartidos $\mathbf{P}_k^{(sh)}$. Para cada uno de estos últimos, también se conoce de forma inmediata cuáles son las fuentes interferentes. Así, es posible escribir:

$$\mathbf{P}_k = \mathbf{P}_k^{(iso)} \bigcup \mathbf{P}_k^{(sh)} \tag{4.25}$$

Partiendo de la información de los parciales aislados y empleando el algoritmo de detección de onsets presentado en la Sección 4.5 [23], es fácil detectar el instante de inicio y el final de cada nota presente. Esta información es muy importante para evitar los artefactos y/o ruido causado por el proceso de mezcla que tiende a aparecer antes y después de las notas activas, el cual resulta acústicamente molesto además de empeorar los resultados numéricos de calidad de la separación.

Considérese un parcial mezclado P_m de frecuencia media $\overline{f_m}$. Este parcial puede escribirse

como:

$$\begin{aligned} P_m(t) = A_m(t)e^{j[\phi_m(t)]} &= \sum_{s_k} P_{s_k}(t) \\ &= \sum_{s_k} A_{s_k}(t)e^{j[\phi_{s_k}(t)]} \end{aligned} \tag{4.26}$$

donde $P_{s_k}(t)$ son los armónicos originales (de las fuentes aisladas) que se solapan. En la Ecuación (4.26), la única información accesible es la amplitud y la fase instantáneas del parcial mezcla, es decir, $A_m(t)$ y $\phi_m(t)$. El objetivo es recuperar cada $A_{s_k}(t)$ y $\phi_{s_k}(t)$ con tanta precisión como sea posible.

Es necesario por lo tanto seleccionar un parcial perteneciente a cada fuente interferente s_k si se desean separar las contribuciones a P_m. Partiendo de los conjuntos de parciales aislados $\mathbf{P}_k^{(iso)}$ correspondientes a cada fuente, se busca un parcial j (para cada fuente) con una energía E_j tan parecida a la energía de P_m como sea posible, y con una frecuencia media $\overline{f_j}$ tan cercana a $\overline{f_m}$ como se pueda. Si $\Delta(E_{j,m}) = |E_j - E_m|$ y $\Delta(f_{j,m}) = |\overline{f_j} - \overline{f_m}|$, estas condiciones pueden escribirse como:

$$P_{k,win} = \left\{ P_j \in \mathbf{P}_k^{(iso)} \quad | \quad \Delta(E_{j,m})|_{min} \right\} \tag{4.27}$$

y:

$$P_{k,win} = \left\{ P_j \in \mathbf{P}_k^{(iso)} \quad | \quad \Delta(f_{j,m})|_{min} \right\} \tag{4.28}$$

La condición de energía, Ecuación (4.27), se calcula en primer lugar. Sólo en casos dudosos la condición frecuencial de la Ecuación (4.28) es evaluada. No obstante, ambas condiciones suelen conducir al mismo parcial ganador y en cualquier caso no proporcionan resultados muy diferentes. Por propósitos de simplicidad, denotemos como P_{wk} al parcial seleccionado (ganador) para cada fuente k. Se puede escribir:

$$P_{wk}(t) = A_{wk}(t)e^{j[\phi_{wk}(t)]} \quad \forall k \tag{4.29}$$

Si $\overline{f_{wk}}$ es la frecuencia media del parcial ganador de la fuente k, es fácil ver que:

$$\frac{\overline{f_{wk}}}{\overline{f_m}} = \frac{p_k}{q_k} \tag{4.30}$$

para ciertos p_k, $q_k \in \mathbb{N}$.

De hecho, el mismo coeficiente p_k/q_k puede ser empleado para reconstruir las frecuencias instantáneas correspondientes a cada fuente interferente con gran precisión, como se verá en el ejemplo detallado de la Sección 4.7.5. A partir de la Ecuación (4.24) se obtienen las fases

instantáneas ϕ_{s_k} correspondientes a las contribuciones de cada fuente[3].

Por otro lado, necesitamos encontrar la mejor combinación lineal de las amplitudes $A_{wk}(t)$ que minimizan el error en la obtención de la amplitud destino (mezclada) $A_m(t)$, es decir:

$$A_m(t_i) = \sum_{s_k} \alpha_k A_{wk}(t_i) \quad \forall t_i \tag{4.31}$$

La Ecuación (4.31) es equivalente a la solución por mínimos cuadrados en presencia de covariancia conocida del sistema:

$$\mathbf{A} * \alpha = b \tag{4.32}$$

donde $\mathbf{A}$ es una matriz que contiene las envolventes de cada parcial seleccionado (ganador), descrito por las Ecuaciones (4.27) y (4.28), α es el vector mezcla y $b = A_m(t)$.

Calculados α_k, p_k y q_k para cada fuente k, el parcial superpuesto es:

$$P_m(t) = \sum_{s_k} \alpha_k A_{wk}(t) e^{j\left[\frac{p_k}{q_k}\phi_{wk}(t)\right]} \tag{4.33}$$

y las contribuciones separadas para cada fuente presente son, por supuesto:

$$P_{s_k}(t) = \alpha_k A_{wk}(t) e^{j\left[\frac{p_k}{q_k}\phi_{wk}(t)\right]} \tag{4.34}$$

Una vez que cada parcial separado se obtiene empleando la técnica descrita, se suma a su correspondiente fuente. Este proceso iterativo conduce eventualmente a las fuentes separadas.

4.7.4. Características generales

A excepción de la inevitable información interferente que pueda estar incluida en los propios parciales aislados (fundamentales y armónicos), la cual está muy diluida bajo la información de la fuente más energética, no existe en general información erróneamente asignada a ninguna fuente (ya que en realidad la información de los parciales mezcla no vuelve a ser utilizada; sólo se emplea para estimar la combinación lineal óptima de proporcionalidades entre las fuentes mezcladas, α). Esto significa que los términos de interferencia en el proceso de separación descrito serán en general despreciables. Por otro lado, en la reconstrucción se tiende intrínsecamente a generar artefactos y distorsiones (ya que las envolventes y fases utilizadas en el proceso están correladas con los datos esperados pero no son, en general, coincidentes). Estas tendencias se verán confirmadas numéricamente en la Sección 4.7.6.

[3] *Supondremos $\Delta\phi_0 = 0$ en la Ecuación (4.24), pero de hecho se puede insertar una fase aleatoria sin que se produzcan diferencias significativas ni en los resultados numéricos ni en los acústicos.*

Las ventajas del proceso descrito son principalmente dos: la primera, los cálculos asociados a la separación de parciales superpuestos (selección de parciales ganadores, cálculo de la combinación lineal óptima por mínimos cuadrados, reconstrucción de las fuentes) no resultan computacionalmente muy pesados. De hecho, la obtención de los propios coeficientes wavelet y en mayor medida la de los parciales, levantados partiendo de máscaras bidimensionales de gran tamaño emplea mucho más tiempo de cálculo. La segunda ventaja es que la separación es completamente ciega. No se necesita *ninguna* característica a priori de las señales de entrada (ni contornos de pitch, ni distribuciones de energía, ni número de fuentes).

4.7.5. Ejemplo detallado

A continuación se va a desarrollar en detalle un ejemplo concreto para clarificar el proceso de separación. Se ha escogido arbitrariamente una de las señales analizadas (ver Sección 4.7.6), compuesta por la mezcla de un vibrato de trompeta $D5$ (587Hz) y una nota $C5$ (523Hz) de trombón tenor. La Ecuación (4.8) de mezcla monaural genérica queda en este caso reducida a:

$$x(t) = s_1(t) + s_2(t) \tag{4.35}$$

La forma de onda, espectrograma wavelet y escalograma de la señal pueden verse en la Figura 4.12, mientras que los resultados numéricos de la separación serán expuestos en la siguiente sección.

Este ejemplo concreto se centra en el parcial compartido más energético. Se han extraído de las señales aisladas los parciales que acaban superponiéndose para generar éste, extrayendo sus características de evolución temporal en amplitud, fase y frecuencia instantánea. Esta información se empleará para demostrar la robustez del método propuesto. Las principales características del parcial superpuesto y de los parciales aislados aparecen más adelante en la Tabla 4.2. En la Tabla 4.3 se reflejan los resultados más relevantes del proceso de separación.

Los resultados exactos del algoritmo de estimación de frecuencias fundamentales son $f_{01} = 589.25$Hz para la trompeta y $f_{02} = 525.96$Hz para el trombón. La amplitud instantánea de los parciales asociados a tales frecuencias se muestra en la Figura 4.13. La linea azul se corresponde con la amplitud del parcial fundamental de la trompeta y la verde con la del trombón tenor.

Tras el análisis armónico, algunos de los parciales detectados serán armónicos naturales de una u otra fuente, generándose de este modo los conjuntos de clase, en este caso $\mathbf{P}_1^{(iso)}$ y $\mathbf{P}_2^{(iso)}$, Ecuación (4.25). Las amplitudes instantáneas de los parciales incluidos en estos conjuntos para el caso presente han sido representadas en la Figura 4.14. En la figura se puede comprobar que los parciales fundamentales forman parte de $\mathbf{P}_{1,2}^{(iso)}$. Sus amplitudes

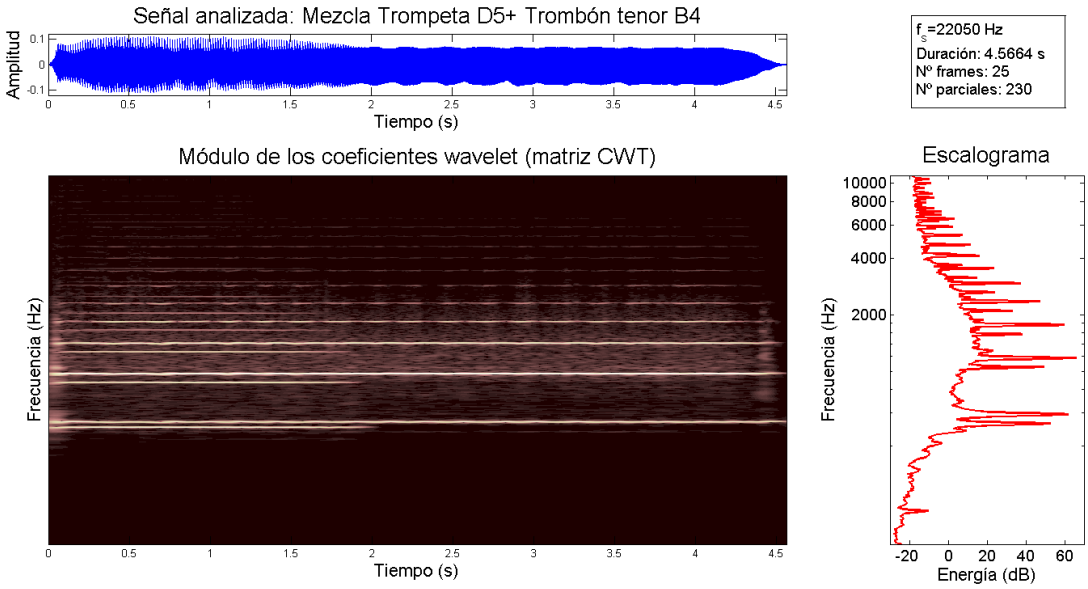


Figura 4.12: *Espectrograma wavelet, escalograma y datos de análisis de la señal mezcla del ejemplo.*

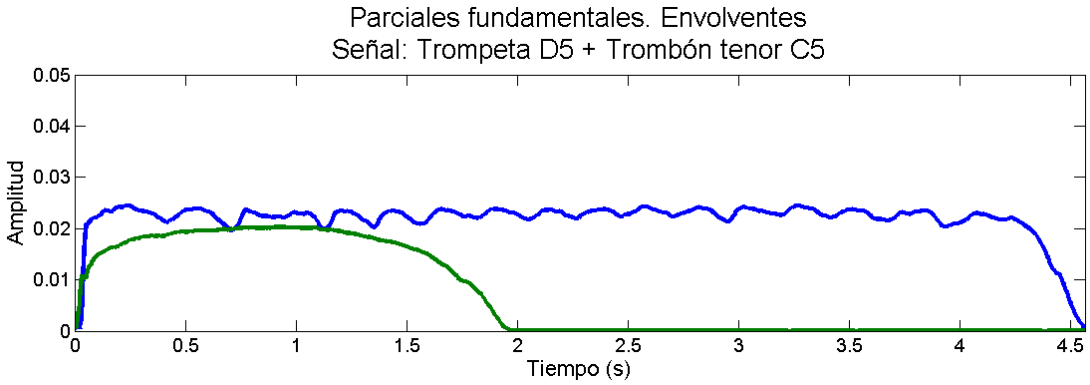


Figura 4.13: *Envolventes de los parciales fundamentales. La línea azul se corresponde con la fundamental de la trompeta. La línea verde, con la del trombón tenor.*

han sido destacadas con líneas más gruesas que el resto.

Durante la separación se procesan uno por uno los parciales superpuestos, de los cuales se conoce además cuáles son las fuentes originales cuyos espectros se solapan en la zona frecuencial correspondiente al parcial mezcla. Partiendo de los conjuntos de parciales aislados de cada fuente, $\mathbf{P}_1^{(iso)}$ y $\mathbf{P}_2^{(iso)}$, se selecciona un parcial para reconstruir perteneciente a cada

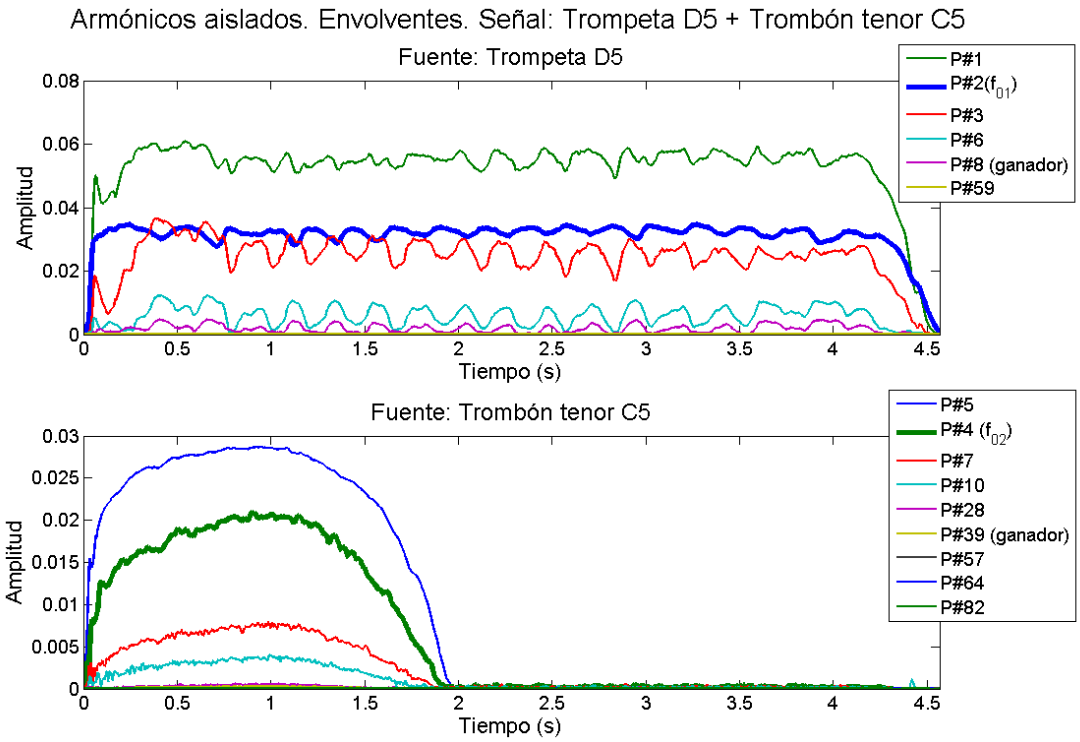


Figura 4.14: *Envolventes de los parciales aislados clasificados para cada fuente. Las envolventes de las fundamentales estén marcadas con líneas gruesas. En azul, la trompeta. En verde, el trombón tenor.*

fuente mezclada. Los parciales escogidos en este caso están indicados en la Figura 4.14 (mediante las etiquetas #8 y #39 respectivamente). Utilizando las frecuencias fundamentales f_{01} y f_{02}, así como la frecuencia media del parcial mezcla, los coeficientes de proporcionalidad $(p_1/q_1, p_2/q_2)$ se calculan mediante la Ecuación (4.24). Estos coeficientes proporcionan la posibilidad de obtener la frecuencia instantánea estimada. El la Figura 4.15 se muestran tanto las frecuencias instantáneas correspondientes a los parciales de las señales aisladas (en azul) como las estimadas a través de este método (en rojo). Como se puede apreciar, el nivel de aproximación en la reconstrucción frecuencial es muy elevado.

El siguiente paso consiste en calcular la solución a la expresión de mínimos cuadrados dada por la Ecuación (4.32). De este modo se busca la combinación lineal óptima de los parciales ganadores que mejor encaja con los datos de amplitud del parcial mezcla. La amplitud instantánea así escalada de cada parcial ganador será considerada como la envolvente de la contribución de su fuente correspondiente. De este modo, los parciales separados serán caracterizados por la solución a la Ecuación (4.31) para amplitudes, y a la Ecuación (4.24)

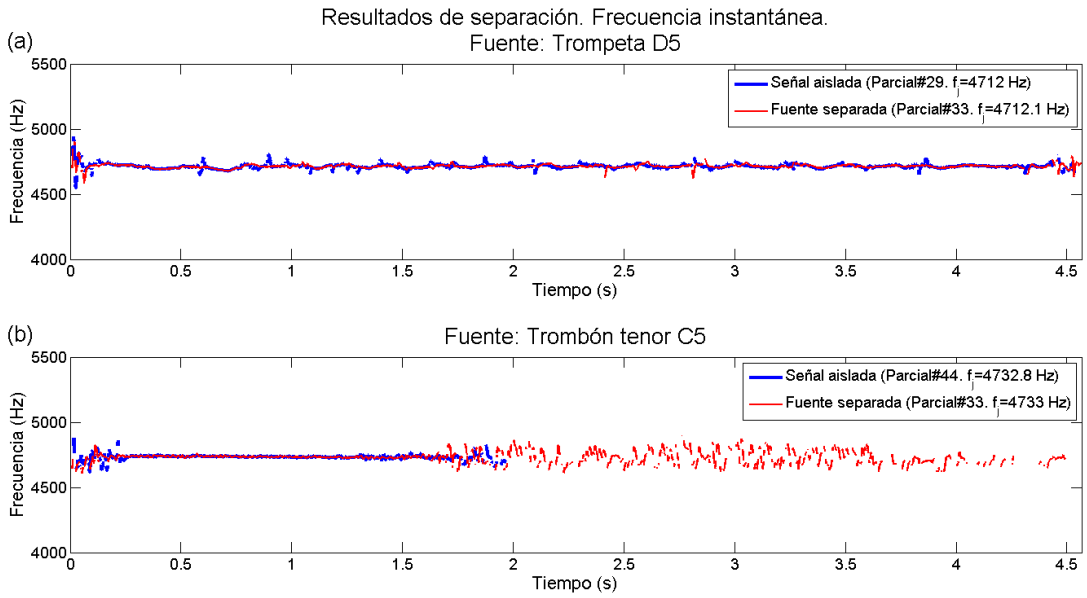


Figura 4.15: *Comparativa entre las frecuencias instantáneas de los parciales originales (aislados) y de los estimados (separados). (a) Resultados para la fuente correspondiente a la trompeta (en azul, la f_{ins} original. En rojo, la estimada). (b) Resultados relativos al trombón tenor.*

para las fases.

En las Tablas 4.2 y 4.3, se detallan algunos de los datos más relevantes del parcial mezcla escogido para este ejemplo de separación.

Datos globales		**Datos del parcial**		
Señal	f_0 (Hz)	$\overline{f_j}$ (Hz)	ϕ_0 (rad)	**RMS** **(dB)**
Trompeta D5	589.27	4712	2.97	-78.46
Trombón C5	525.94	4732.8	-0.86	-87.55

Tabla 4.2: *Datos principales de las señales originales (aisladas) y del parcial superpuesto del ejemplo.*

Concretamente, en la Tabla 4.2, se muestran los datos correspondientes a las señales aisladas, que incluyen las frecuencias fundamentales, así como las frecuencias medias de los parciales que generan el parcial compartido del ejemplo y sus correspondientes fases inicia-

Datos globales de la señal mezcla		**Datos del parcial separado**				
Fuente	f_0 (Hz)	$\overline{f_j}$ (Hz)	p	q	ϕ_0 (rad)	**RMS (dB)**
$s_1(D5)$	589.25	4712.1	8	9	-2.36	-79.78
$s_2(C5)$	525.96	4733	5	7	-2.3	-94.68

Tabla 4.3: *Datos principales de la señal mezclada y de las contribuciones separadas obtenidas para el parcial superpuesto del ejemplo.*

les (información obtenida a través del algoritmo CWAS), y los valores RMS que indican la energía de tales parciales. En la Tabla 4.3, se han reflejado los datos numéricos correspondientes al proceso de separación, incluyendo las frecuencias fundamentales detectadas en la señal mezcla, las frecuencias medias de los parciales separados (partiendo del parcial mezcla del ejemplo), los valores experimentales obtenidos para los parámetros de proporcionalidad p y q para cada fuente, las fases iniciales de los parciales reconstruidos y los valores RMS correspondientes a las contribuciones separadas. Estas tablas sirven de apoyo a la precisión en la separación (corroborada en los resultados numéricos de la Sección 4.7.6) tanto como para demostrar la dificultad del problema al que nos enfrentamos (obsérvese la discrepancia entre los valores de las fases iniciales reales y separadas).

En la Figura 4.16 aparecen en azul las formas de onda de los parciales originales que se superponen en el parcial mezcla, obtenidas a través del análisis wavelet de las señales aisladas. En rojo, las contribuciones separadas obtenidas para cada fuente.

Una vez obtenido cada parcial separado a través de este método, estos son añadidos a la fuente adecuada. El proceso iterativo proporciona eventualmente las fuentes separadas. En la Figura 4.17 se han representado las formas de onda correspondientes al proceso de separación detallado. En la primera de las gráficas, Figura 4.17 (a), la señal mezcla original (trompeta $D5$ más trombón tenor $C5$) En las Figuras 4.17 (b) y (d), las señales aisladas originales (en azul). Por último, en las Figuras 4.17 (c) y (e) las fuentes separadas (en rojo).

En la Figura 4.18 se muestran los espectros de Fourier de las diferentes señales implicadas. El primer espectro, Figura 4.18 (a), se corresponde con la señal mezcla. En las gráficas siguientes, Figuras 4.18 (b) y (c), los espectros de la trompeta y el trombón tenor, respectivamente. En cada una de estas gráficas, se ha representado por el trazo azul el espectro de la señal original, y en rojo el espectro correspondiente a la fuente separada.

Como se puede observar en los espectros, la mayoría de la parte armónica de cada fuente ha sido adecuadamente separada. Los parciales incorrectamente estimados no suelen

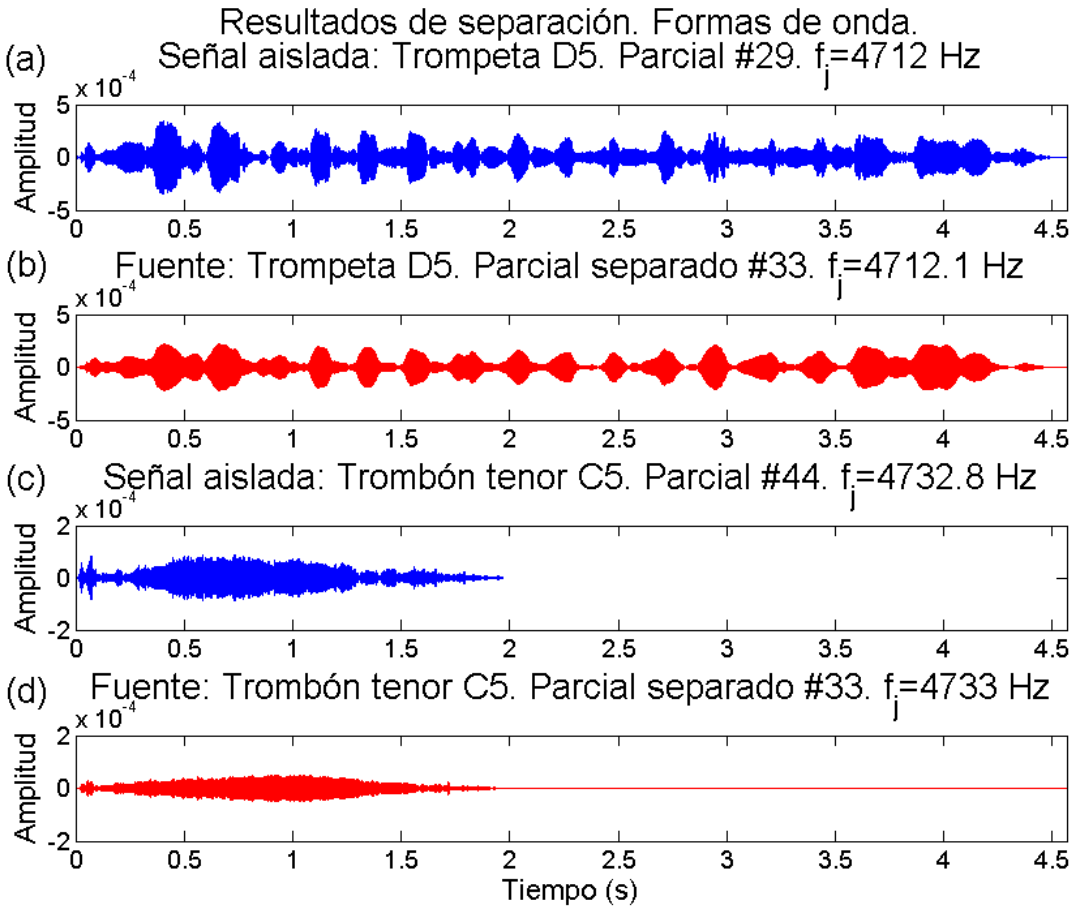


Figura 4.16: *(a) y (c): Formas de onda de los parciales originales. En azul, trompeta y trombón tenor, respectivamente. (b) y (d): Formas de onda de las contribuciones separadas. En rojo, trompeta y trombón tenor, respectivamente.*

tener, en general, energía suficiente para redundar en un error relevante. Sin embargo, en ocasiones, bien los errores temporales en las envolventes de los parciales de alta energía o la no inclusión de la parte no armónica de las fuentes en el modelo pueden desembocar en diferencias tímbricas perceptibles.

Como conclusiones generales del proceso descrito, las frecuencias fundamentales son localizadas con gran precisión, así como las frecuencias instantáneas de los parciales superpuestos. La energía de estos parciales se recupera con un grado de aproximación suficiente como para poder reconstruir una señal que presenta muy pocas diferencias sonoras respecto a la señal original. Sin embargo, las fases iniciales de los parciales aislados y las obteni-

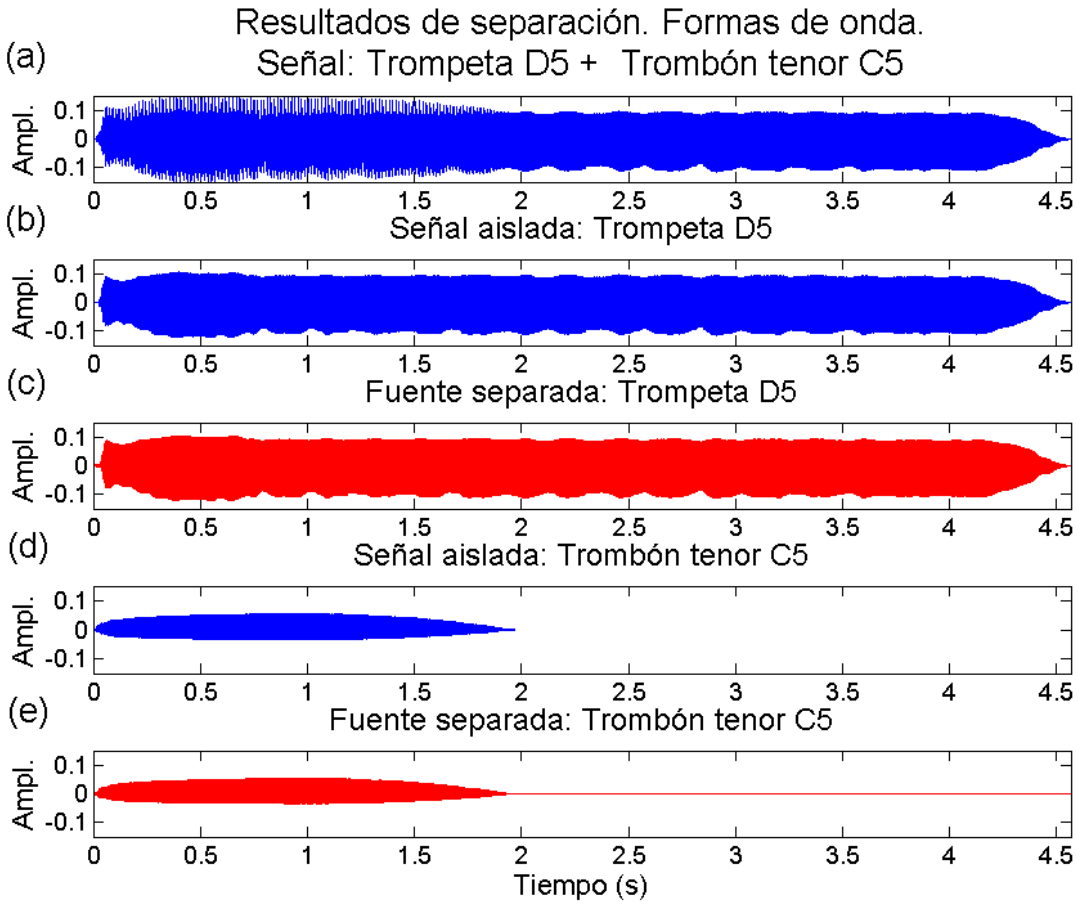


Figura 4.17: *Formas de onda finales. (a) Señal mezcla. (b) y (c) Trompeta original (azul) y separada (rojo). (d) y (e) Trombón tenor original (azul) y separado (rojo).*

das tras el proceso de separación son obviamente diferentes. Esto último es la causa por la cual la técnica propuesta no es a priori capaz de explotar la potencialidad intrínseca del algoritmo CWAS a la hora de obtener las formas de onda exactas de los parciales aislados y por lo tanto de las fuentes separadas; tan solo se recupera la estructura general de las mismas, la forma de su envolvente. Esto no supone en general una gran desventaja, puesto que si las envolventes son suficientemente parecidas y las evoluciones frecuenciales poseen un alto grado de semejanza, los parciales separados tienden a sonar de forma muy similar a los originales, y por lo tanto las fuentes separadas retienen la mayoría de las características tímbricas de las señales aisladas (o de su parte armónica, para ser exactos).

En las Figuras 4.19 y 4.20 aparecen representados los espectrogramas wavelet correspon-

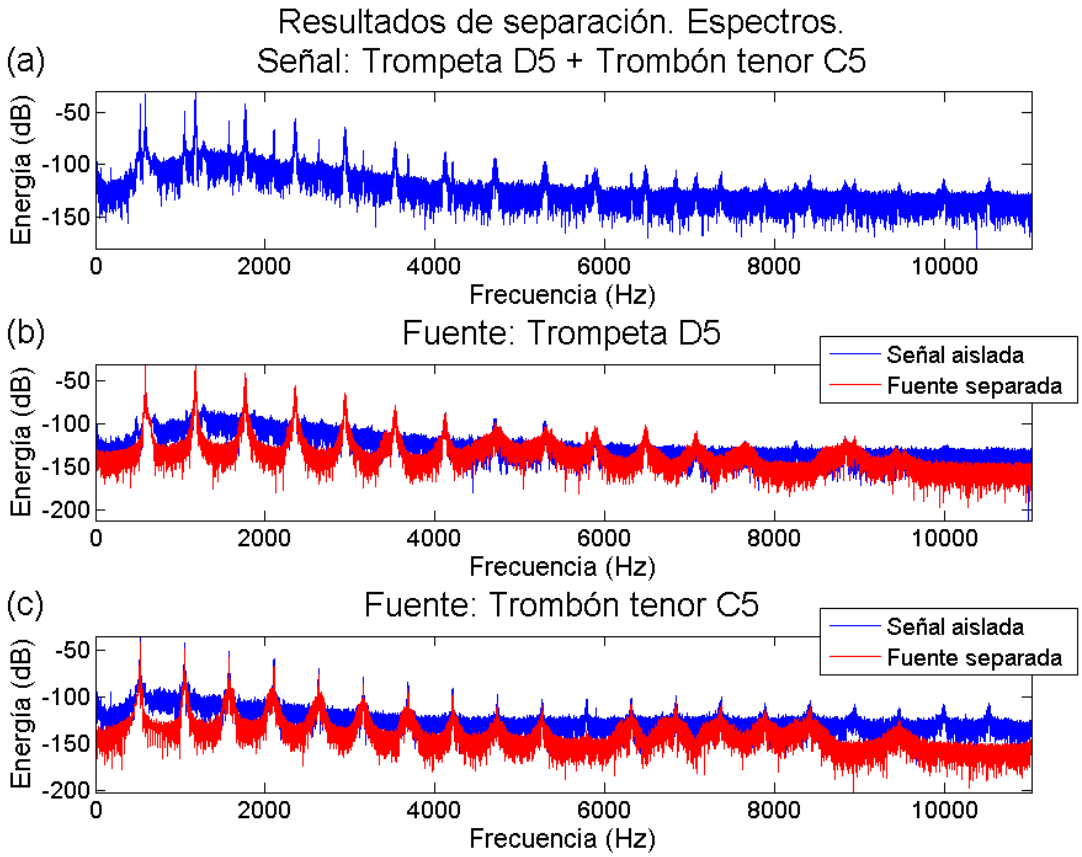


Figura 4.18: *(a) Espectro de la señal mezcla. (b) Espectros de la trompeta aislada (azul) y de la separada (rojo). (c) Espectros del trombón tenor aislado (azul) y separado (rojo).*

dientes con la señal original (izquierda) y la fuente separada (derecha) de trombón tenor y trompeta, respectivamente. En el centro de la figura, el escalograma de cada señal aislada (en azul) y de su correspondiente fuente separada (en rojo). Una vez más, ahora a partir de los escalogramas, se deduce que la información armónica de las señales ha sido adecuadamente reconstruida. Sin embargo, en los espectrogramas se puede ver cómo hay información entre armónicos que no se ha tenido en cuenta en la separación. En las señales en las que esta información sea algo más que simple ruido (por ejemplo, los soplidos especialmente presentes en señales de instrumentos de viento), se notarán diferencias sonoras entre originales y fuentes separadas, como se ha anticipado más arriba.

Los valores de los parámetros estándar de calidad para la señal de este ejemplo se detallan en las Tablas III.9 a III.11 de la siguiente sección.

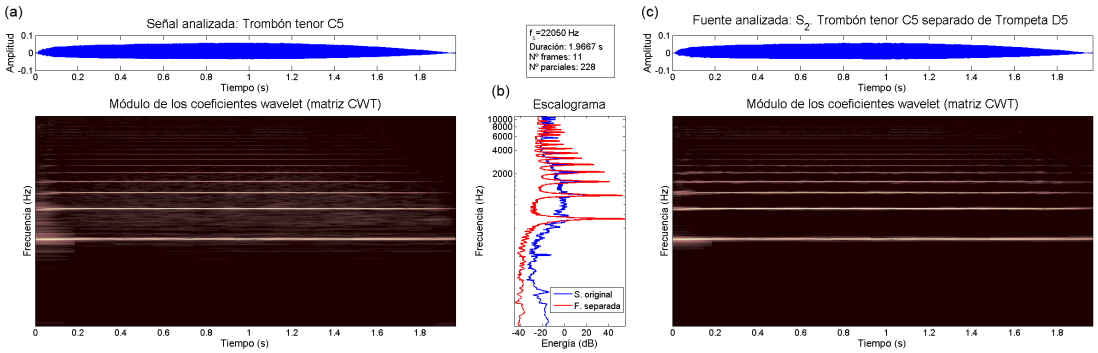


Figura 4.19: *Espectrogramas de las señales del trombón tenor. (a) Espectrograma wavelet del trombón original (aislado). (b) En azul: escalograma original. En rojo: escalograma de la fuente separada. Caja superior: otra información correspondiente a la fuente original. (c) Espectrograma wavelet de la fuente separada.*

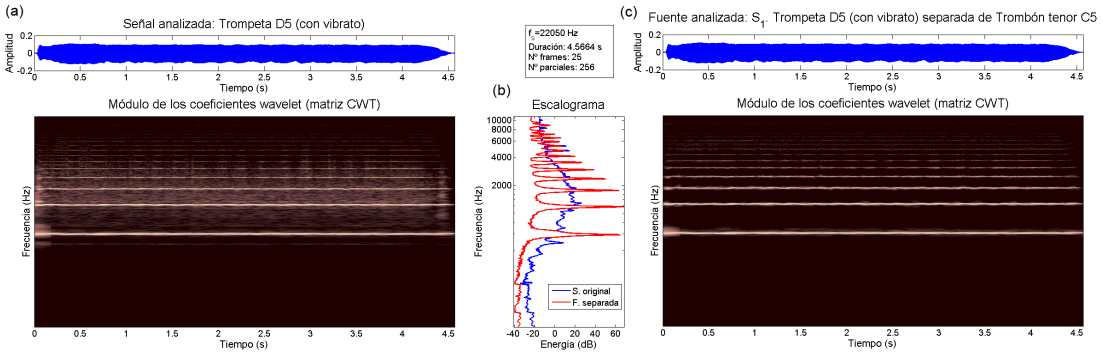


Figura 4.20: *Espectrogramas de las señales de la trompeta. (a) Espectrograma wavelet de la trompeta original (aislada). (b) En azul: escalograma original. En rojo: escalograma de la fuente separada. Caja superior: otra información correspondiente a la fuente original. (c) Espectrograma wavelet de la fuente separada.*

4.7.6. Resultados experimentales

El conjunto de señales analizadas por este procedimiento incluye 56 señales mezcla de 2 fuentes y 15 señales mezcla de 3 fuentes. Todos los instrumentos son grabaciones reales, la mayoría de ellos extraídos de la base de datos de Iowa [63]. El conjunto de instrumentos musicales empleados consta de flauta, clarinete, saxo, fagot, trombón, trompeta, oboe, cuerno, tuba, violín, viola, guitarra y piano.

Todas las señales analizadas han sido subsampleadas a $f_s = 22050$Hz y mezcladas

sintéticamente. En este caso, conviene que el número de parciales aislados sea el más alto posible. Por lo tanto, como se adelantó anteriormente, se ha variado el número de divisiones por octava D del algoritmo, aumentando la resolución hasta un total de 769 bandas de frecuencia (por defecto, en el resto de esta disertación se ha trabajado con 189 bandas, ver Sección 3.4). El vector concreto de divisiones por octava tomado ha sido $D = \{16; 32; 64; 128; 128; 100; 100; 100; 100\}$. Otros parámetros de interés son θ_{th}=0.03, E_{th}=1 %.

4.7.6.1. Pruebas desarrolladas

Se han diseñado un total de 8 tests con 2 y 3 fuentes mezcladas sintéticamente. La lista de experimentos aparece en la Tabla 4.4. En los siguientes párrafos se explicará cada uno de estos ensayos más en detalle. Los resultados gráficos y numéricos se presentarán en la Sección 4.7.6.2.

Experimentos BASS desarrollados			
Experimento (#)	**Fuentes (#)**	**Instrumentos involucrados**	**Características del experimento**
1	2	Diferentes	1 Arm. + 1 Inarm.
2	2	El mismo	Misma octava
3	2	El mismo	Intervalos de 5^a y 12^a
4	2	Diferentes	Intervalos de 5^a y 12^a
5	2	Diferentes	Notas inarmónicas
6	3	El mismo	Acorde mayor
7	3	El mismo	Acorde menor
8	3	Diferentes	Notas inarmónicas

Tabla 4.4: *Lista de experimentos desarrollados.*

Experimento #1: instrumentos armónico e inarmónico

En el primer experimento se han mezclado un instrumento inarmónico (piano) con uno armónico (trompeta, clarinete en *Si bemol* y flauta), generando un total de 3 señales. Los datos numéricos se presentan en la primera columna de las Figuras 4.23 a 4.21. En comparación con los demás resultados, los resultados numéricos de este experimento no son demasiado buenos (acústicamente la situación es mucho mejor), probablemente debido a la incertidumbre en el parámetro de inarmonicidad β [127] (Sección 4.7.1).

Experimento #2: un mismo instrumento, misma octava

En el segundo test, se escogió de forma aleatoria un instrumento musical de entre los disponibles. Partiendo de las notas grabadas de este instrumento (en concreto un saxo *alto*) se han generado un total de 11 señales con las mezclas de dos notas dentro de la cuarta octava (considerando $A4 = 440$Hz). Una de las notas es fija, una $C\#4$ (277Hz), las otras son los 11 tonos restantes dentro de la octava ($C4$, $D4$, $D\#4$, etc.). Los valores experimentales de los parámetros SDR, SIR y SAR se presentan en la segunda columna de las Figuras 4.23 a 4.21.

Experimento #3: un mismo instrumento, notas relacionadas armónicamente

En el tercer ensayo se han mezclado dos notas relacionadas armónicamente ejecutadas por el mismo instrumento. Las relaciones armónicas seleccionadas son: $C - G$, $D - A$, $E - B$, $F - C$, $G - D$, $A - E$ y $A\# - F$, dentro de la misma o en diferentes octavas (es decir, intervalos de 5^a y 12^a). Se han generado dos conjuntos de señales, cada una correspondiente a un instrumento musical (saxo *alto* y clarinete en *Si bemol*), con 9 mezclas para cada uno. Los resultados numéricos se han representado gráficamente en la tercera columna de las Figuras 4.23 a 4.21.

Experimento #4: dos instrumentos, notas relacionadas armónicamente

A continuación se han sintetizado 9 mezclas con las mismas relaciones armónicas del experimento anterior, esta vez ejecutadas por dos instrumentos musicales diferentes: saxo alto, guitarra, fagot, clarinetes en *Mi* y en *Si bemol*, cuerno y flauta. Los valores experimentales de los parámetros de medida en calidad de la separación se presentan en la cuarta columna de las Figuras 4.23 a 4.21.

Experimento #5: dos instrumentos, notas no relacionadas armónicamente

En este experimento se han generado 15 señales, cada una conteniendo la mezcla de dos instrumentos musicales escogidos aleatoriamente tocando notas no relacionadas armónicamente. Los valores numéricos de los parámetros SDR, SIR y SAR aparecen en la quinta columna de las Figuras 4.23 a 4.21.

Experimento #6: mismo instrumento, acorde mayor

Un acorde mayor es la mezcla de tres notas, en concreto $C - E - G$. Se han generado un total de 5 de estos acordes tocados por el mismo instrumento musical (fagot, saxo alto, clarinete en *Si bemol*, flauta y trompeta). Los datos numéricos se presentan en las Figuras 4.23 a 4.21, sexta columna.

Experimento #7: mismo instrumento, acorde menor

Un acorde menor es la mezcla de las notas $A - C - E$. Se han analizado un total de 5 señales, cada una de ellas interpretada por un mismo instrumento musical (fagot, clarinete en *Si bemol*, cuerno, oboe y trompeta). Los resultados de los parámetros SDR, SIR y SAR aparecen en la séptima columna de las Figuras 4.23 a 4.21.

Experimento #8: tres instrumentos, notas no relacionadas armónicamente

Por último, 5 señales con tres instrumentos musicales tocando notas aleatorias (no relacionadas armónicamente). Los valores de calidad en la separación aparecen en la última columna de las Figuras 4.23 a 4.21.

4.7.6.2. Resultados

Los resultados numéricos correspondientes a los 8 experimentos detallados aparecen en las Figuras 4.23 a 4.21. En concreto, en la Figura 4.23 se presentan los resultados del parámetro SDR , en la Figura 4.22, los correspondientes al SIR y finalmente en la Figura 4.21, se muestran los datos numéricos del parámetro SAR.

En la Figura 4.23, los resultados de calidad para cada test aparecen marcados con cuadrados. Con triángulos, los valores extremos obtenidos en cada prueba. Estos resultados corroboran que la separación de notas armónicamente relacionadas es el experimento más difícil de los llevados a cabo con 2 fuentes. En el caso de 3 fuentes, las diferencias no son significativas.

En la Figura 4.22, los resultados numéricos del parámetro SIR aparecen marcados con círculos. De nuevo se delimitan con triángulos los valores máximo y mínimo en cada test. Como puede verse en la figura, los valores numéricos del parámetro SIR presentan una variabilidad menor que en el caso previo, lo que significa que la técnica propuesta no presenta una tendencia significativa a verse afectada por términos de interferencia.

Por último, en la Figura 4.21, los resultados del parámetro SAR para cada prueba están marcados con estrellas. Este parámetro presenta valores muy similares al SDR, por lo que las conclusiones son las mismas.

Considerando globalmente las mezclas de 2 fuentes se obtienen los correspondientes valores promedio de cada parámetro, que aparecen representados en las Figuras 4.23 a 4.21 mediante líneas de trazos y puntos. Los correspondientes valores numéricos son:

- $\overline{SDR_{2s}} \approx 16.98$ dB.
- $\overline{SIR_{2s}} \approx 59.61$ dB.
- $\overline{SAR_{2s}} \approx 17.00$ dB.

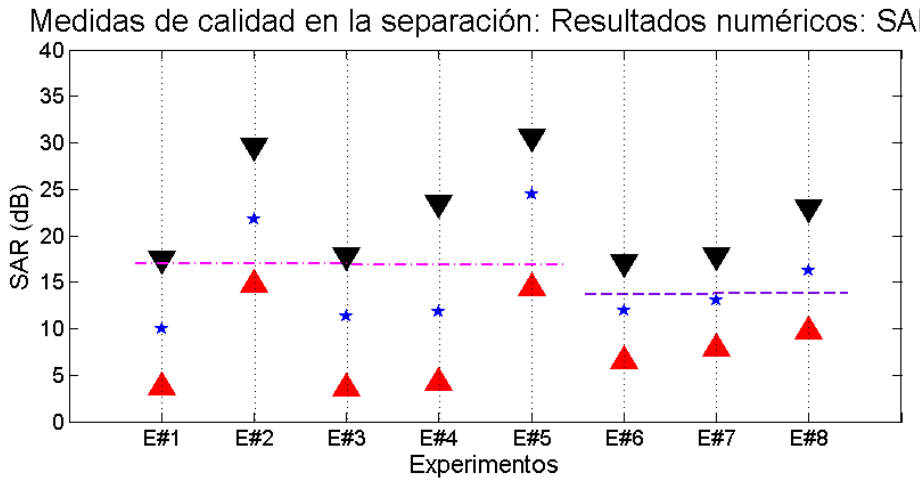


Figura 4.21: *Resultados numéricos en calidad de la separación para cada uno de los 8 experimentos realizados: parámetro SAR.*

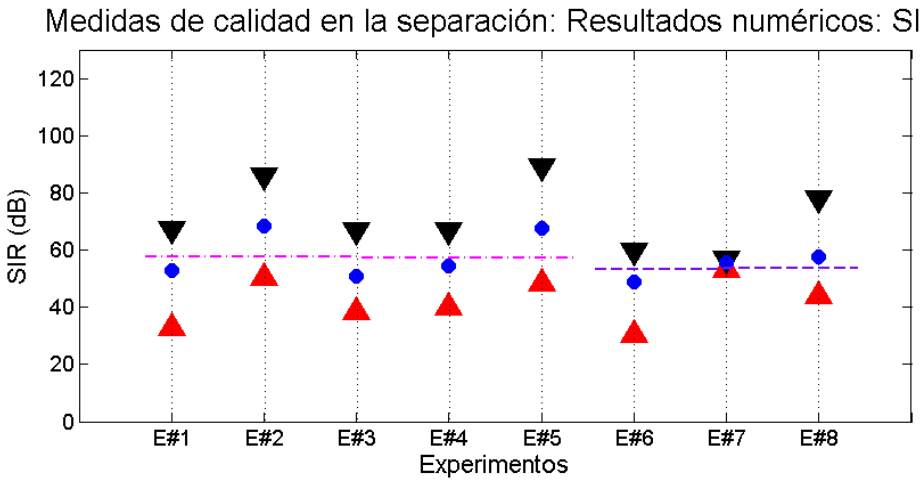


Figura 4.22: *Resultados numéricos en calidad de la separación para cada uno de los 8 experimentos realizados: parámetro SIR.*

Para el caso de las mezclas de 3 fuentes (líneas horizontales discontinuas en las Figuras 4.23 a 4.21), los valores correspondientes son:

- $\overline{SDR_{3s}} \approx 13.78$ dB.
- $\overline{SIR_{3s}} \approx 53.99$ dB.

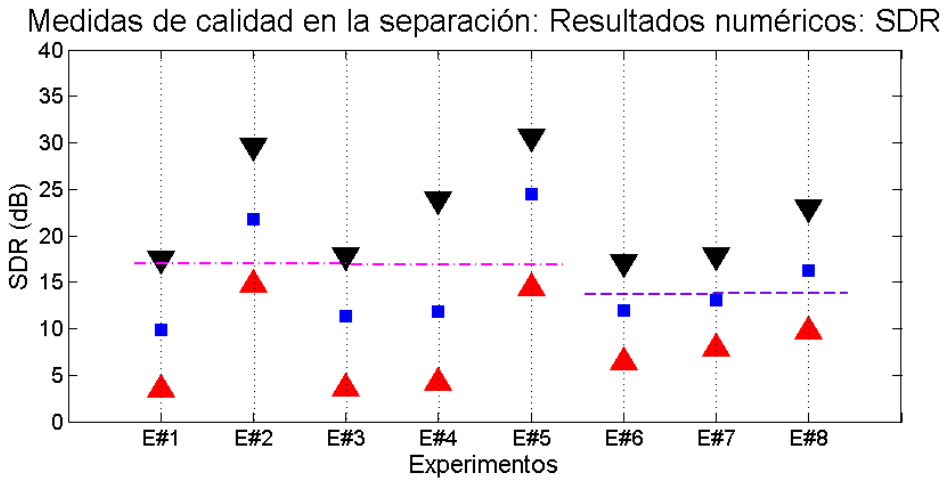


Figura 4.23: *Resultados numéricos en calidad de la separación para cada uno de los 8 experimentos realizados: parámetro SDR.*

- $\overline{SAR_{3s}} \approx 13.79$ dB.

Manteniendo la precisión frecuencial constante, a medida que se incrementa el número de fuentes la separación entre parciales decrece y por lo tanto la amplitud de las interferencias entre fuentes tiende a incrementarse, lo que se refleja en el valor del parámetro SIR. El número de parciales aislados es igualmente menor, por lo que los artefactos y distorsiones tienden asimismo a crecer, afectando a los parámetros SDR y SAR.

4.7.6.3. Limitaciones y valoración

Esta última técnica de separación ciega de fuentes presenta dos diferencias principales respecto a otras existentes: en primer lugar, evidentemente, la herramienta utilizada para llevar a cabo el análisis tiempo–frecuencia, en nuestro caso el algoritmo CWAS. La alta coherencia temporal y frecuencial obtenida en los parciales aislados supone un punto ventajoso de partida con respecto, por ejemplo, al análisis STFT estándar. Por otro lado, los parciales separados se reconstruyen completamente partiendo de los aislados, buscando la mejor combinación lineal de amplitudes instantáneas que minimiza el error en la envolvente generada en el parcial mezcla, asumiendo el principio de CAM. Al utilizar en el proceso de reconstrucción parciales con energía similar (tanto como permitan las opciones disponibles en las clases aisladas) se tiende a mejorar el resultado final del proceso. Es decir, si el parcial superpuesto tiene alta energía, se empleará en la reconstrucción parciales aislados de alta energía, con lo que los valores de correlación se espera que sean asimismo elevados [110, 173].

Si por el contrario el parcial mezcla posee una energía baja, se utilizarán parciales de energía baja cuya correlación con los ideales tal vez sea menor, pero que tienden a dotar a la señal de un color muy similar al presente en la fuente original y causan errores generalmente asumibles. De este modo, aunque las distorsiones y artificios tiendan a resultar más evidentes que las interferencias, los resultados sonoros resultan ser sensiblemente mejores al menos respecto a los presentados en las otras técnicas del presente capítulo. De hecho, la presencia de la/s fuentes/s interferente/s suele resultar acústicamente despreciable. De este modo, la reconstrucción de fase no es tan importante como en otras técnicas disponibles en la literatura, pudiéndose obtener fuentes separadas que presentan a la vez resultados numéricos de calidad muy elevados y alta semejanza sonora con respecto a las señales originales. Sin embargo, la técnica aquí desarrollada, en su estado actual, es capaz de separar exclusivamente notas de instrumentos musicales armónicos de forma global, es decir, no se incluyen en el modelo sonidos inarmónicos ni señales en las que el mismo instrumento musical ejecute más de una nota. Un objetivo a corto plazo sería implementar el algoritmo de separación en un contexto frame-to-frame, con lo cual se obtendría un separador de notas muy eficiente.

4.8. Evolución futura

Ciertos instrumentos musicales quedan marcados por la información no armónica que los acompaña (soplidos o otros ruidos de ejecución). Una posible mejora en la calidad final de la separación sería la inclusión de la información no armónica detectada en la mezcla, lo que podría dar como resultado una recuperación más eficiente del timbre. El problema de fondo es cómo asignar esta información a cada fuente de forma correcta y automática.

Independientemente de este punto, es evidente que el siguiente tema a abordar es el de la separación de temas musicales más complejos, con varios instrumentos diferentes ejecutando distintas notas. Ya que en este momento el método de separación ciega de notas musicales es un buen separador de notas genérico, cabe esperar que los resultados de tal separación sean asimismo de alta calidad.

En la mayor parte de los trabajos consultados se recurre a la partitura de la mezcla para poder asignar cada nota separada a su correspondiente fuente (lo cual descarta la obtención automática de partituras como aplicación). Este objetivo parece alcanzable a corto plazo, al menos en primera aproximación (ya que en un tema musical es muy probable que, por ejemplo, aparezcan notas *completamente superpuestas* correspondientes a distintos instrumentos, posibilidad no contemplada por el método propuesto, aunque abordable, como se ha explicado en la Sección 4.2.3).

Existen formas alternativas de lograr la separación sin recurrir a la partitura. Una de ellas es el empleo de modelos de instrumentos (los cuales pueden generarse por ejemplo mediante redes neuronales). En este tipo de procesos, la separación tiende a especializarse

en un conjunto de instrumentos musicales bastante reducido. Otra posibilidad a tener en cuenta consiste en realizar un análisis tímbrico de las diferentes fuentes presentes, intentando encontrar (por ejemplo mediante la envolvente *cepstral*) qué conjunto de las notas detectadas se corresponde con cada una de ellas.

Los avances que se han obtenido en la separación de fuentes musicales pueden afinarse mediante la aplicación de modelos de excitación-resonancia para aplicaciones orientadas a la voz, de modo que las posibilidades de evolución futura y/o especialización de la técnica resultan evidentes.

Por descontado, esta es una de las líneas de investigación principales en este momento del Grupo de Audio Digital de la Universidad de Zaragoza. Se sigue avanzando en la generalización del proceso presentado, en la búsqueda de aplicaciones más prácticas del mismo, en el desarrollo de un algoritmo (*ad hoc*, si es necesario) para la separación y mejora de la voz hablada o cantada, así como en aplicar técnicas de separación multicanal (en concreto estéreo mediante DUET) al algoritmo CWAS.

4.9. Conclusiones y contribuciones

En el presente Capítulo se ha presentado una técnica de separación de notas armónicas en señales musicales monaurales basada en el algoritmo CWAS. Esta técnica necesita a su vez dos bloques modulares adicionales, en concreto un algoritmo de localización de onsets y un estimador de frecuencias fundamentales para señales multipitch.

En cuanto a las contribuciones presentadas en este Capítulo, cabe destacar que todos los algoritmos detallados son de creación propia. Éstos se han diseñado de forma específica para las diferentes aplicaciones, si bien algunas de sus partes pueden haber sido adaptadas de técnicas preexistentes.

Concretando, las contribuciones principales son:

1. Algoritmo de localización de onsets.

2. Algoritmos de detección de frecuencia fundamental (2).

3. Algoritmo de separación ciega de fuentes de audio monaurales por reconstrucción de parciales superpuestos.

   Previamente a esta técnica de separación, se han desarrollado dos métodos alternativos de separación de sonidos monaurales (detallados en el Anexo III):

   - Por onsets.
   - Por distancias.

Además, se han desarrollado aplicaciones adicionales que no han sido incluidas en el cuerpo principal de esta Tesis por motivos de extensión. Tales aplicaciones serán presentadas en el Anexo II, y entre ellas caben destacar principalmente las siguientes:

4. Algoritmo de análisis sub-banda.

5. Algoritmo simple de filtrado de señales.

6. Algoritmos de efectos musicales (17).

Una de las líneas de investigación que permanecen abiertas es la separación de fuentes de sonido en grabaciones monocanal.

Capítulo 5

Comparativas del Algoritmo C.W.A.S.

Índice

"Si buscas resultados distintos, no hagas siempre lo mismo".
Albert Einstein (1879-1955).
Físico de origen alemán,
nacionalizado suizo y estadounidense.

En este último capítulo se establece un balance entre el algoritmo desarrollado y otros modelos y técnicas existentes en cuatro aspectos: el tiempo de computación requerido, en la extracción de características de la señal de audio, en la representación visual (y el acceso a la información) y para finalizar, en la calidad final de la resíntesis.

5.1. Introducción

A lo largo de los dos capítulos anteriores se ha demostrado que el algoritmo CWAS presenta una muy elevada precisión en la obtención de parámetros de alto nivel de la señal de audio. Es el momento de evaluar diferentes características de la técnica propuesta en comparación con otras herramientas existentes, lo cual se va a efectuar en cuatro aspectos distintos:

- Tiempo de proceso.

  Como se ha detallado ampliamente, el algoritmo CWAS ofrece, muestra a muestra, el valor tanto de la amplitud instantánea (envolvente) como de la fase y frecuencia instantáneas de cada parcial detectado en la señal de entrada. El tiempo de procesado que se requiere para acceder a los coeficientes wavelet (y por lo tanto a la información característica de la señal) es relativamente alto. Sin embargo, esta lentitud es tan sólo aparente. Para comparar de un modo serio los tiempos de computación necesarios debe ponerse a trabajar a la FFT en condiciones equivalentes (en cuanto a precisión) a las del algoritmo CWAS. De esta forma, se analizarán un total de 25 señales de audio cuyas duraciones oscilan entre los 500 milisegundos y más de 36 segundos. Cada una de estas señales ha sido analizada ocho veces, cinco sobre un ordenador de sobremesa de prestaciones elevadas y tres sobre un notebook, sensiblemente menos potente, en ambos equipos bajo entorno Matlab®, v7.8.0.347. El propósito de éste experimento es que sus resultados puedan servir para extraer conclusiones de velocidad de proceso en relación a la duración de la señal (e independientemente de otras características particulares de la misma), atendiendo además a la máquina en que ésta haya sido analizada.

- Obtención de frecuencia instantánea.

  En este caso, se compararán sendas herramientas de análisis con el algoritmo CWAS a la hora de obtener la frecuencia instantánea de señales de audio: la *Time-Frequency Toolbox* de Matlab®, desarrollada por François Auger y otros [7, 10, 11], (empleada con permiso de los autores) y las rutinas espectrográficas de alta resolución, desarrolladas por Sean Fulop [64]. Empleando ocho de los algoritmos disponibles en tales herramientas, se ha analizado un conjunto ocho señales sintéticas. Se comparará el valor teórico de cada ley frecuencial generada con los datos experimentales obtenidos empleando estas técnicas, tres de ellas basadas en la STFT, otras tres basadas en la WVD, el algoritmo de Fulop y nuestro algoritmo CWAS.

- Representación visual de la información.

  En donde se compararán las diferentes técnicas de representación visual de la evolución temporal del espectro de la señal de audio que se han desarrollado, con otras representaciones tiempo-frecuencia. Concretamente, se comparará el espectrograma wavelet con el espectrograma estándar (empleando los datos obtenidos en el análisis de tiempos de computación así como como el escalograma obtenido a través del algoritmo de Fulop y la Time-Frequency Toolbox de Auger, STFT y WVD). Por último, se comparará la representación bidimensional y tridimensional de los parciales del modelo propuesto con la conocida técnica de Reasignación [92], con la intención de comparar las diferencias no sólo en la calidad visual de la información, sino en sus demás características inherentes.

- Precisión en la resíntesis.

  Para finalizar, se van a presentar los resultados obtenidos con el algoritmo CWAS en la generación de sonidos sintéticos. Como se demostrará, estos resultados son de altísima calidad, hasta el punto de que las diferencias numéricas y acústicas entre la señal analizada (independientemente de su naturaleza y duración) y la señal sintética resultan despreciables tanto numérica como acústicamente. Los resultados de nuestro algoritmo se compararán con los obtenidos mediante el uso de la *Síntesis por Modelado eSprectral* (SMS) [153], desarrollado por Xavier Serra y su grupo [148].

5.2. Tiempo de computación

La Transformada de Fourier Localizada o Transformada Corta de Fourier, STFT (presentada en la Sección 1.4.2), dado un tamaño de ventana de análisis (*window size*, w, el cual se suele tomar impar para que el centro de la ventana esté localizado en un valor entero), y un tamaño de superposición de ventanas (*overlap*, o), obtiene una muestra de análisis cada $w - o$ muestras (tamaño de salto, o *hopsize*), la cual queda posicionada en la malla T–F

sobre el centro de la ventana de análisis. Como se explicó en la citada Sección 1.4.2, para obtener la información de las $w - o - 1$ muestras restantes, se interpolan los datos obtenidos (por ejemplo mediante un algoritmo de interpolación bicúbica).

Si se pretende obtener la información de la STFT punto por punto, es necesario que el tamaño de superposición o sea 1 muestra menor que el de la ventana de análisis w. En este caso, se computará la transformada rápida de Fourier FFT (un algoritmo de sobras conocido y ampliamente optimizado), una vez por cada muestra de la señal.

Por ejemplo, una señal de medio segundo de duración, muestreada a $f_s = 44100$Hz, necesitará de un total de 22050 FFT's para ser analizada en su totalidad. Sin embargo, el algoritmo CWAS empleará una FFT y una iFFT por cada banda de análisis (de cardinal ligado al número de divisiones por octava D) para obtener, en un sólo paso, los resultados punto por punto de 4095 muestras de señal analizada. Dado que, salvo en el caso de la separación de fuentes, se ha trabajado con un máximo de 201 bandas de análisis wavelet (para $f_s = 44100$Hz, 189 para $f_s = 22050$Hz), el número de operaciones necesarias para obtener resultados comparables entre ambos métodos parece ser ampliamente favorable al algoritmo CWAS (para la señal de medio segundo citada anteriormente, 2412 FFT's en total: dos por cada banda frecuencial en cada uno de los seis frames de análisis, frente a 22050 FFT's necesarias en la STFT, lo cual indica una rapidez teórica de cálculo 9.14 veces superior para el algoritmo CWAS). Este resultado aproximado no es nuevo. De hecho, la transformada wavelet en general es intrínsecamente más rápida que la STFT bajo cualesquiera idénticas condiciones de trabajo [120]. Los resultados, que se mostrarán en la Sección 5.2.3, indican la misma conclusión cualitativa.

5.2.1. La Transformada de Fourier Localizada, muestra a muestra

Se ha programado un algoritmo de análisis de señales de audio basado en la STFT que refleja la información tempo–frecuencial de la señal de entrada con un *hopsize* de 1. Los parámetros más relevantes del análisis son:

- Tamaño de ventana (winsize)[1]: 2047.
- Tamaño de salto (hopsize): 1.
- Resolución frecuencial: 2048.

[1] *El tamaño de ventana está ligado a la capacidad de la FFT para distinguir transitorios. A mayor tamaño de ventana, más período de tiempo estudiado globalmente y por lo tanto menor capacidad de distinguir eventos de corta duración. Ventanas más pequeñas permiten localizar con mayor precisión los transitorios, pero afectan a la capacidad de la FFT de sacar partido a su longitud de bins frecuenciales. El valor de* window size *escogido es,* grosso modo*, un compromiso entre precisión temporal y frecuencial.*

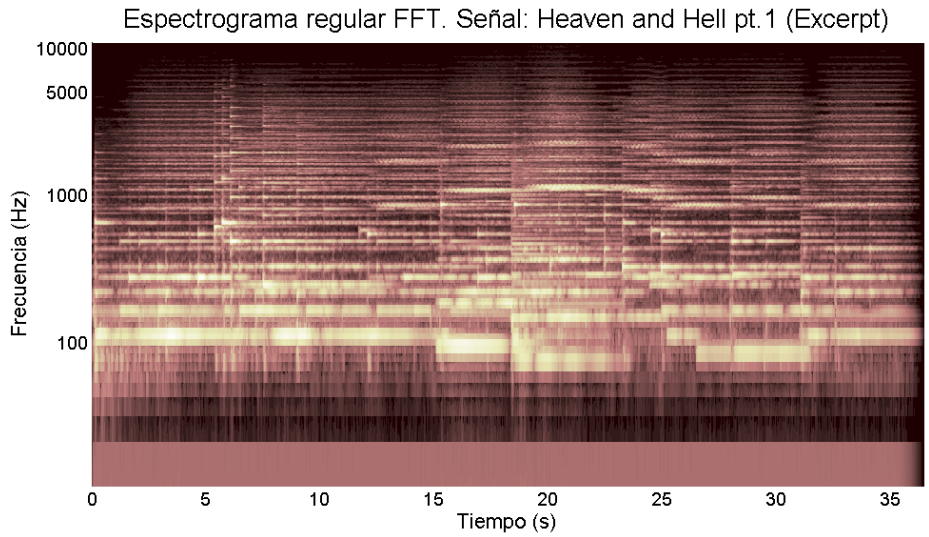


Figura 5.1: *Ejemplo de un espectrograma regular FFT. La señal analizada es un extracto de 36 segundos del conocido tema de Vangelis* "Heaven and Hell, Part 1".

A partir del módulo de la matriz de coeficientes de Fourier, se puede representar gráficamente la variación temporal del espectro de la señal, en un espectrograma al que se llamará en adelante *espectrograma regular* FFT, para distinguirlo del espectrograma wavelet que se ha utilizado hasta ahora.

En la Figura 5.1 aparece el espectrograma regular FFT de una señal de audio (concretamente, un extracto de 36 segundos aproximadamente, del tercer movimiento del tema *"Heaven and Hell, Part 1"*, de Vangelis)[2], analizado bajo los parámetros anteriormente descritos. Cada uno de los 2048 bins frecuenciales se encuentra equiespaciado en el eje de frecuencias. La representación semilogarítmica de la figura causa un ensanchamiento de los filtros situados en la parte baja del espectro, diseminando parcialmente la información frecuencial en la parte inferior de la gráfica, y condensándola en la superior.

5.2.2. Acerca del algoritmo CWAS

Por otro lado, el algoritmo CWAS presenta la capacidad de poder escoger diferentes divisiones en cada octava del espectro (algo no programado *a priori* en la FFT), con lo cual se permite un control exhaustivo de la capacidad resolutiva del algoritmo, muy útil en determinadas ocasiones. El algoritmo CWAS ha sido utilizado, para el presente estudio comparativo, bajo las consideradas condiciones estándar de trabajo, que son:

[2] *"Heaven and Hell, Part 1", procedente del álbum "Heaven and Hell", de Vangelis. ©RCA Ltd., 1975.*

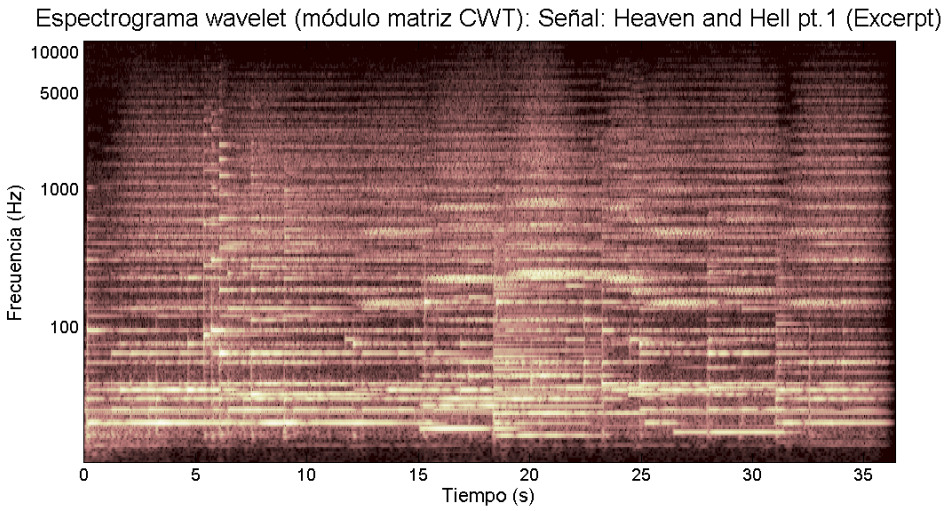


Figura 5.2: *Espectrograma wavelet correspondiente a la misma señal representada en la Figura 5.1.*

- Tamaño de frame: 4095.
- Resolución frecuencial: 201 bandas para $f_s = 44100$Hz, 189 para $f_s = 22050$Hz.

En la Figura 5.2 se ha representado el espectrograma wavelet de la misma señal *"Heaven and Hell, part 1"* representada en la Figura 5.1. Como se puede observar, ambas gráficas son similares, si bien el módulo de los coeficientes wavelet resulta ser más revelador en la zona de baja frecuencia, a cambio de perder capacidad de separación en la zona de alta, por los motivos expuestos anteriormente. El algoritmo CWAS posee aproximadamente 10 veces menos bins frecuenciales que la STFT programada.

5.2.3. Comparativa de tiempos de computación

Se ha analizado un conjunto de 25 señales de diferentes duraciones por ambos métodos. Cada una de éstas señales ha sido analizada cinco veces en un equipo de sobremesa *MEDION*, Intel® Core™i7, CPU 920 @ 2.67GHz, con 12GB de RAM y sistema operativo Windows7 Enterprise™64bits, y tres veces más en un notebook *ASUS EeePC*, Intel® Atom™, CPU N450 @ 1.66GHz, con 1GB de RAM y sistema operativo Windows7 Starter™32bits. Analizando cada señal varias veces, se pretende que los procesos internos del sistema operativo influyan lo menos posible en el resultado final. De este modo se establece un marco comparativo eficiente entre la STFT y el algoritmo CWAS, sólo dependiente de la máquina (es decir, básicamente de la cantidad de RAM disponible y, en menor medida, de la velocidad del microprocesador).

Los resultados experimentales obtenidos se encuentran resumidos en la Tabla 5.1, donde aparece, para cada computador empleado, el tiempo promedio del análisis de Fourier y el del análisis Wavelet, así como la relación entre ambos tiempos, calculada como sigue:

$$Ratio = \frac{t_{FFT}}{t_{CWT}} \tag{5.1}$$

Como se puede ver en la tabla, el algoritmo CWAS es sensiblemente más rápido que la STFT en ambos equipos y con todas las señales. Parece haber una relación entre la duración de la señal y el ratio de tiempos, resultando éste más favorable al algoritmo CWAS a medida que la duración de la señal aumenta (el coeficiente de tiempos de computación en el equipo MEDION es de 24.44 para la señal s_{13}, de 36 segundos de duración).

Tomando el tiempo total empleado en el análisis de Fourier y el empleado en el análisis Wavelet, se obtienen los siguientes coeficientes de relación promedio:

- $Ratio_1 = 14.26\ (9.19)$[3].
- $Ratio_2 = 6.40\ (6.04)$[4].

Para una información más detallada acerca de los tiempos de computación obtenidos, así como de las señales analizadas, consúltese el Anexo IV.a.

5.3. Recuperación de la frecuencia instantánea

Una vez despejada la cuestión del tiempo de procesado, se va a proceder a comparar la capacidad de extracción de las leyes frecuenciales presentes en la señal de audio. Para ello, se emplearán, como se ha avanzado anteriormente, tanto las rutinas espectrográficas de alta resolución (HRSR) de Sean Fulop[5] [58, 64], como la Time-Frequency Toolbox (TFTB), de François Auger[6] [7], ambas desarrolladas en entorno Matlab®. Concretando un poco más, en cuanto a los scripts de Fulop, se calculará el espectro de potencia de Nelson [121, 122]. Respecto al trabajo de Auger, se emplearán más sus algoritmos de cálculo de la STFT y la WVD Suavizada, incluyendo las versiones reasignadas y los algoritmos de extracción de crestas (ridges) y esqueletos.

[3] *Equipo de sobremesa* MEDION*, Intel® Core™i7, CPU 920 @ 2.67GHz, 12GB RAM, sistema operativo Windows7 Enterprise™64bits. El número entre paréntesis indica el resultado del ratio de tiempos excluyendo los resultados de la señal s_{13}, "Heaven and Hell Pt.1 (Excerpt)", de coeficiente especialmente favorable al algoritmo CWAS.*

[4] *Notebook* ASUS EeePC*, Intel® Atom™, CPU N450 @ 1.66GHz, 1GB RAM, sistema operativo Windows7 Starter™32bits. El número entre paréntesis tiene la misma explicación que en el caso anterior.*

[5] *Las HRSR pueden descargarse desde la web de MatlabCentral, en MathWorks [115]. La URL de descarga es:* `http://www.mathworks.com/matlabcentral/fileexchange/21736-high-resolution-spectrographic-routines`.

[6] *La TFTB ha sido desarrollada principalmente bajo los auspicios del CNRS (*Centre National de la Recherche Scientifique*) francés. Parte de los scripts ha sido desarrollada en la Universidad de Rice (Inglaterra). Disponible en varias páginas web, entre otras:* `http://tftb.nongnu.org`.

		Tiempo de procesado					
		Intel® Core™i7 @ 2.67GHz 12GB RAM, OS. 64bits			Intel® Atom™@ 1.66GHz 1GB RAM, OS. 32bits		
Señal	**L(m)**	$\overline{t_{FFT}}$**(s)**	$\overline{t_{CWAS}}$**(s)**	**Ratio**	$\overline{t_{FFT}}$**(s)**	$\overline{t_{CWAS}}$**(s)**	**Ratio**
s_1	61423	94.2654	8.9581	10.5231	459.7605	62.8732	7.3074
s_2	58553	72.5401	9.1060	7.9668	278.9440	52.5965	5.3037
s_3	57776	88.3179	8.9794	9.8364	365.5397	57.5208	6.3588
s_4	49606	75.2876	7.8245	9.6221	307.0420	48.3646	6.3486
s_5	25228	30.3619	4.4847	6.7706	129.7546	28.1464	4.6100
s_6	140390	229.9736	20.3890	11.2798	869.1052	123.3435	7.0465
s_7	76598	96.1848	11.1964	8.5909	362.4195	64.6337	5.6073
s_8	44084	53.8850	6.7320	8.0044	226.7668	41.5035	5.4640
s_9	97732	125.7651	14.0273	8.9661	500.9798	85.5603	5.8554
s_{10}	20755	24.6423	3.9510	6.2394	106.0013	24.0828	4.4019
s_{11}	45189	55.4138	7.3562	7.5343	233.0967	45.0143	5.1786
s_{12}	34881	42.0732	5.5545	7.5748	177.8766	34.0721	5.2206
s_{13}	803133	2706.4800	110.7379	24.4411	4531.4242	633.2453	7.1560
s_{14}	66150	102.1350	10.0945	10.1180	410.9289	64.5711	6.3710
s_{15}	61583	76.6738	9.5280	8.0473	321.1211	58.7482	5.4662
s_{16}	38956	47.1186	6.1135	7.7074	197.6844	37.3516	5.2926
s_{17}	22051	26.2124	4.6648	5.6192	108.8213	28.9275	3.7645
s_{18}	55929	83.9644	8.4304	9.9600	344.8164	51.7968	6.6572
s_{19}	57330	87.8136	8.5356	10.2901	323.4312	48.9648	6.6055
s_{20}	105840	168.1933	15.2199	11.0517	652.3250	92.6280	7.0425
s_{21}	60809	93.4899	8.9900	10.3995	342.0585	53.1806	6.4393
s_{22}	50840	76.0293	7.8441	9.6926	313.7000	48.5434	6.4626
s_{23}	83057	105.4827	12.4042	8.5047	435.7953	75.7352	5.7548
s_{24}	100412	130.2351	14.6795	8.8720	525.8091	89.3946	5.8821
s_{25}	46514	56.7886	7.3477	7.7288	244.1215	45.6095	5.3537

Tabla 5.1: *Datos de tiempo de computación promedio para cada una de las 25 señales analizadas, y en cada uno de los dos computadores empleados. Estos ordenadores son: Equipo de sobremesa* MEDION*, Intel® Core™i7, CPU 920 @ 2.67GHz, 12GB RAM, sistema operativo Windows7 Enterprise™64bits, y notebook* ASUS EeePC*, Intel® Atom™, CPU N450 @ 1.66GHz, 1GB RAM, sistema operativo Windows7 Starter™32bits.*

5.3.1. Time-Frequency Toolbox

La TFTB [7, 10, 11] es una colección de aproximadamente 100 scripts para GNU Octave y Matlab® desarrollados para el análisis de señales no estacionarias utilizando transformaciones tiempo–frecuencia. Está dirigida principalmente a investigadores, ingenieros y

estudiantes con conocimientos básicos en procesamiento de la señal. Ha sido desarrollada para Matlab® v4.2c y posteriores (aunque ha sido necesario modificar ligeramente algunos scripts para hacerlos compatibles con la última version empleada, v7.0.4), y requiere la *Signal Processing Toolbox* v3.0 o posterior [7].

Aunque la TFTB incluye más de una docena de TFD, en las siguientes Secciones se emplearán básicamente dos de ellas, en concreto la STFT y la Distribución Pseudo Wigner-Ville, o Wigner-Ville Suavizada (PWVD), además del Reassignment de cada una de ellas, y sus representaciones cresta-esqueleto.

Esta herramienta fue desarrollada a finales de la década de los '90, por lo que no incluye algunas de las reformas más significativas que se han llevado a cabo en el análisis de señales no estacionarias. Los scripts utilizados manejan un volumen de información tal, que los parámetros de control del análisis deben ser relativamente pobres (como se verá en las siguientes secciones) para evitar problemas de memoria. Pese a todo, el proceso completo a través del cual se generan los datos de Reasignación puede llevar varios minutos.

5.3.1.1. Espectrograma STFT

La STFT ha sido presentada con cierto detalle en la Sección 1.4.2. En lo que se refiere a la TFTB, la función que calcula el espectrograma STFT (módulo cuadrático de la STFT), dada una señal $x(t)$ discreta, tiene varios parámetros de control importantes, cuyos valores para el presente análisis son los siguientes:

- Muestras de $x(t)$ analizadas: señal completa.
- Número de bins frecuenciales: 2048.
- Ventana de análisis $h(t)$: *Hanning* de 2049 muestras.

5.3.1.2. Distribución Pseudo Wigner-Ville

La Distribución de Wigner-Ville ha sido a su vez presentada en la Sección 1.4.3. Como se dijo en su momento, presenta una larga lista de propiedades matemáticas deseables, entre las que cabe destacar que es definida real y covariante ante desplazamientos temporales y frecuenciales.

A diferencia de los espectrogramas regular y wavelet, los términos de interferencia de la WVD son no nulos sin importar la distancia entre los términos interferentes de la señal, generándose valores espurios de la distribución en lugares donde la respuesta debería ser cero. Estos términos de interferencia pueden suponer un problema, ya que eventualmente podrían superponerse con términos espectrales procedentes de la señal, lo cual hace difícil interpretar visualmente la imagen de la WVD. Sin embargo, como se apuntó en la Sección

1.4.3, parece que estos términos deben permanecer presentes, o las propiedades matemáticas de la WVD pueden dejar de satisfacerse. En la TFTB, se trabaja con un consenso entre la presencia de interferencias y las propiedades satisfechas.

La PWVD se define matemáticamente como sigue:

$$PWVD_x(k,\phi) = \int_{-\infty}^{+\infty} h(\tau)x(k+\frac{\tau}{2})x^*(k-\frac{\tau}{2})e^{-j\phi\tau}d\tau \tag{5.2}$$

donde $h(t)$ es una ventana estándar, cuyo objetivo es el suavizado frecuencial de la distribución [7]. Debido a la naturaleza oscilatoria de los términos de interferencia, éstos se verán atenuados mediante el uso de esta ventana, la cual causa, como se ha dicho, la potencial pérdida de algunas propiedades matemáticas de la WVD, así como una mayor apertura en frecuencia de las componentes propias de la señal.

En el análisis llevado a cabo, la función que calcula los coeficientes de la PWVD está controlada por los mismos parámetros que en el caso del espectrograma regular STFT, y los valores empleados de tales parámetros han sido:

- Muestras de $x(t)$ analizadas: señal completa.
- Número de bins frecuenciales: 2048.
- Ventana de análisis $h(t)$: *Hanning* de 2049 muestras.

5.3.1.3. Reassignment

La falta de resolución inherente a cualquier distribución tiempo–frecuencia (debida al ancho de banda instantáneo) se puede remediar parcialmente con el *espectrograma de frecuencia instantánea corregida en tiempo* (TCIFS, Time-Corrected Instantaneous Frequency Spectrogram), más conocido, en inglés, por *reassignment*, y que será referido en adelante como *reasignación*.

En la reasignación, se calculan estimaciones afinadas de las componentes espectrales a partir de las derivadas parciales del espectro de fase [59]. En efecto, en lugar de localizar las componentes en el centro geométrico de la ventana de análisis (como se ha explicado que sucede en el espectrograma regular), las componentes se reasignan a los *centros de gravedad* de la distribución espectral compleja de energía, calculados aplicando el *principio de fase estacionaria* introducido a su vez en la Sección 1.5.3. Éste principio concluye que la energía de la señal (distribuida en el semiplano T–F), contribuye más significativamente en las zonas del espectro donde la fase posee una variación lenta comparada con el entorno. Como se verá más adelante, cada punto del espectrograma se relocaliza en su centro de gravedad asociado más próximo (puntos que pueden tener valores en tiempo y frecuencia no enteros, es decir, fuera de la malla original), obteniéndose de esta forma un espectrograma

relocalizado. Este proceso queda resumido en la Figura 5.3. La TFD original lleva asignada una malla de análisis con una cierta apertura en tiempo y frecuencia, Figura 5.3 (a). Cada punto del análisis con suficiente energía es reasignado siguiendo alguno de los métodos existentes [9, 92, 93, 121, 122], de modo que de una distribución equiespaciada, Figura 5.3 (b), se pasa a una distribución irregular más cercana a la verdadera región de soporte de la señal, Figura 5.3 (c).

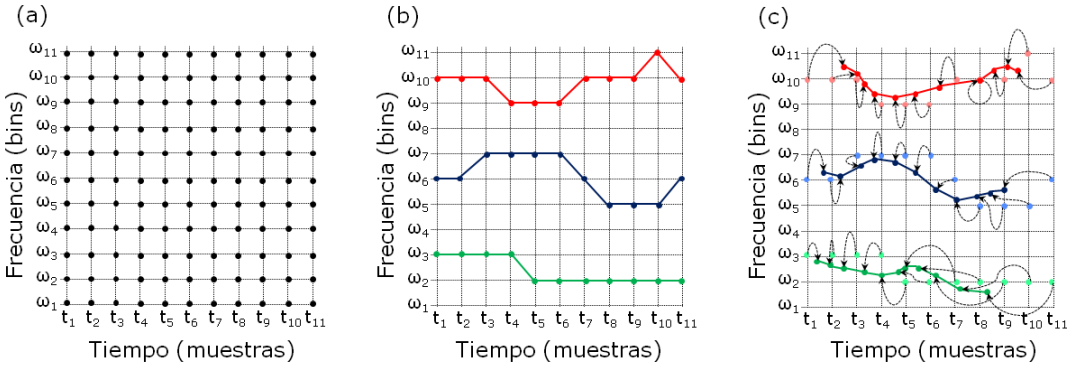


Figura 5.3: *Concepto teórico de la Reasignación. Sobre una malla tiempo frecuencia determinada (a), se localizan las posiciones de concentración energética suficiente (b), situadas sobre puntos concretos de la malla. La Reasignación mueve esos puntos sobre variables continuas, reposicionando las componentes (c).*

5.3.1.3.1. Distribuciones tiempo–frecuencia y reasignación

El origen último de la reasignación es la relación entre la fase de la señal *analítica* y su frecuencia instantánea [65], establecida por Rihaczek [137] (Capítulo 1), si bien no fue hasta 1976 [92] que se propuso el método de reasignación original, llamado Método de Ventana Móvil Modificada (*Modified Moving Window Method*, MMWM), siendo aplicado en concreto a la búsqueda de leyes de variación frecuencial en pulsaciones geomagnéticas. Aunque el MMWM inicial ya proporcionaba resultados mejores que otras técnicas de análisis de señales no estacionarias preexistentes [93], fue probablemente a causa de una base teórica poco sólida que la técnica permaneció prácticamente ignorada por la comunidad científica durante casi dos décadas más, hasta que trabajos posteriores [8, 9, 45] mejoraron esta base y devolvieron al germen de la TCIFS a la investigación internacional activa. Posteriores revisiones [121, 122] han aumentado sensiblemente la precisión de la reasignación [58], lo cual ha llevado directamente a una gran diversificación de sus empleos [59], como el análisis de alta precisión de señales musicales [78], el afinamiento del espectro de señales tipo chirp

[174], la detección de componentes no estacionarias (y en particular de transitorios, difícil de conseguir mediante métodos más tradicionales) [177], o el análisis de señales recogidas por sensores magnetostrictivos y provenientes de motores de DC [61].

En cuanto al método de reasignación matemático, éste puede ser calculado de formas muy diversas [9, 92, 93, 121, 122], cuyos resultados, si bien ligeramente diferentes, son similares. Por concretar un poco más, una de las posibles formas de calcular las posiciones tiempo–frecuencia reasignadas (tomada de [59]) es:

$$\hat{t}_{k,n} = n - \Re\left[\frac{X_{t;n}X_n^*(k)}{|X_n(k)|^2}\right] \tag{5.3}$$

$$\hat{\omega}_{k,n} = k + \Im\left[\frac{X_{f;n}X_n^*(k)}{|X_n(k)|^2}\right] \tag{5.4}$$

donde $\widehat{t}_{k,n}$ y $\widehat{\omega}_{k,n}$ son las posiciones temporal y frecuencial corregidas, correspondientes a la $k - sima$ componente espectral del análisis centrada en el tiempo n (en muestras, asumiendo ventanas de análisis de longitud impar). $X_{t;n}$ y $X_{f;n}$ denotan las transformadas localizadas, computadas empleando una ponderación enventanada temporal y frecuencial (respectivamente) de la señal, mientras que $\Re[\cdot]$ y $\Im[\cdot]$ son la parte real y la imaginaria de lo encerrado entre corchetes. Cabe destacar que tanto $\widehat{t}_{k,n}$ como $\widehat{\omega}_{k,n}$ pueden tener valores de muestra fraccionarios.

5.3.1.4. Crestas (ridges)

Para terminar, dada cualquier TFD, se pueden calcular sus *crestas* y *esqueletos*. Dado que se trata exactamente del mismo concepto presentado en la Sección 1.5.3, no se ahondará en qué consisten matemáticamente. Sus mismos nombres son suficientemente aclaratorios, además del hecho de que se trata de extraer la información mínima que caracteriza la señal, ligada a los máximos locales de ésta en el semiplano T–F.

En el análisis que nos ocupa, se han extraído las localizaciones temporales y frecuenciales que caracterizan las diferentes componentes de la señal, para el caso tanto del espectrograma regular como de la PWVD.

5.3.2. Rutinas espectrográficas de alta resolución

Las *rutinas espectrográficas de alta resolución* (High Resolution Spectrographic Routines, HRSR) [64] de Sean Fulop, son un conjunto de seis pequeños programas desarrollados para Matlab®, que, basándose en el espectro de potencia de Nelson [121, 122], van más allá del análisis de Fourier obteniendo, de forma alternativa a la propuesta por Auger, el espectro reasignado (o de alta resolución) de la señal de entrada. Son, por lo tanto, posteriores al desarrollo de la TFTB, y como se verá más adelante, proporcionan una frecuencia

instantánea canalizada que tiende a mejorar sensiblemente los resultados de los algoritmos de Auger tanto en precisión como sobre todo en tiempo de proceso.

Entre las diferentes posibilidades, se ha empleado el script *"Nelsonspecjet_both.m"*, cuyo fundamento teórico aparece detallado en [58], el cual calcula el espectrograma reasignado a través del algoritmo de espectro cruzado de Nelson [121]. Las HRSR en general proporcionan una salida que consta de cierto número de puntos localizados en las zonas del semiplano T–F donde se acumula la energía de la señal. Estos puntos son de fase estacionaria, como se ha dicho. Al analizar señales monocomponente, la salida del espectro cruzado de Nelson muestra un claro agrupamiento de estos puntos alrededor de los datos reales de frecuencia instantánea. Sin embargo, en señales multicomponente aparecen además un gran número de puntos que se distribuyen de forma aleatoria, lo que dificulta bastante la obtención de los datos de frecuencia instantánea canalizada correspondientes a cada componente. Para ilustrar este hecho, se incluye la Figura 5.4, correspondiente a los resultados del análisis de la señal Mezcla de tres sinusoidales de frecuencias 400Hz, 600Hz y 800Hz.

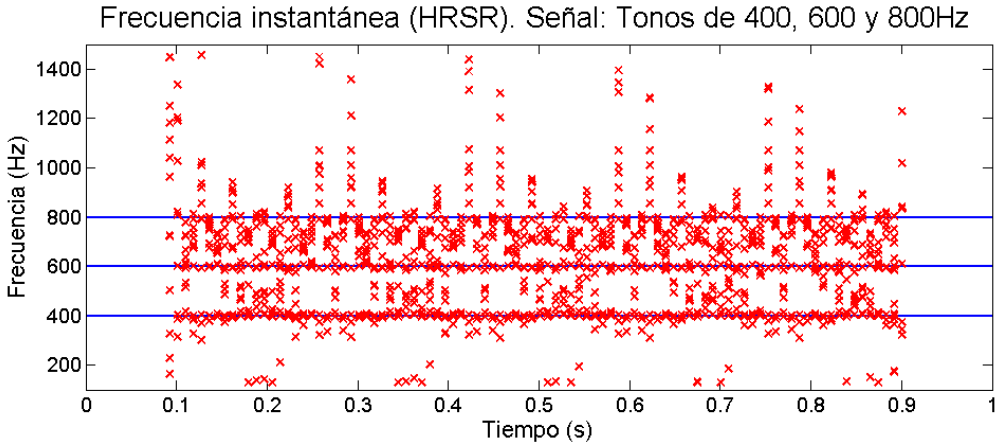


Figura 5.4: *Salida HRSR. Señal: mezcla de tres tonos puros de frecuencias 400, 600 y 800Hz. En azul, las frecuencias instantáneas teóricas. Con cruces rojas aparecen los datos de frecuencia instantánea canalizada.*

Esta distribución de los datos nos ha obligado a programar un algoritmo de *peak picking* que localiza los puntos de mayor concentración de energía en el escalograma de la señal y busca los puntos del semiplano responsables de tal acumulación, separándolos en los datos de las correspondientes componentes, con lo que se pueden obtener los datos de frecuencia instantánea de cada una de ellas.

En cuanto al script empleado (así como los demás incluidos en las HRSR) presenta algunas características que conviene detallar brevemente. En primer lugar, al estar basado en la STFT, necesita de un tamaño de ventana de análisis (winsize) y de una resolución frecuencial determinada (en bins). El parámetro más importante es el tamaño de ventana, puesto

que, como se ha explicado, supone el verdadero límite resolutivo del algoritmo. Ventanas muy pequeñas localizarán muy bien eventos repentinos, pero causarán un espectro muy poco resolutivo en frecuencia, y viceversa. Para tener una resolución frecuencial adecuada, se hace necesario emplear ventanas relativamente grandes. Sin embargo, por el diseño interno del algoritmo, ventanas grandes causan un número de puntos temporales donde se conoce la frecuencia cada vez más pequeño (situación poco recomendable para señales sintéticas complicadas y grabaciones reales). Por lo tanto, se hace necesario un compromiso entre la ventana de análisis y el número final de puntos.

Los parámetros de control con los que se ha trabajado, han sido:

- Número de bins frecuenciales: 2048.
- Ventana de análisis $h(t)$: *Hanning*[7] de 255 ó 511 muestras, dependiendo de la señal.
- Overlap: 128 muestras.
- Frecuencias inferior y superior de corte: 20Hz y $f_s/2$, respectivamente.
- Clip[8]: -96dB.

5.3.3. Recuperación de frecuencia instantánea: comparativa

Para comparar la precisión en la recuperación de la frecuencia instantánea, se ha procedido al análisis de un total de ocho señales sintéticas, cuyas características más importantes se resumen en la Tabla 5.2.

Cuatro de estas señales serán utilizadas más adelante (en concreto, la señal de FM con excursión sinusoidal, y los chirps lineal, cuadrático e hiperbólico, Sección 5.5.1). Hay una quinta señal con idénticas duración y frecuencia de muestreo, concretamente un tono puro de 440Hz. Se han generado tres señales nuevas ex profeso para esta comparativa, todas ellas de 1 segundo de duración y $f_s = 44100$Hz: Una mezcla ponderada de tres tonos de 400Hz, 600Hz y 800Hz (de pesos relativos 1, 0.7 y 0.3 respectivamente), un chirp cuya frecuencia aumenta hiperbólicamente entre los 100Hz y los 8kHz el primer medio segundo y desciende de 8kHz a 100Hz linealmente en el siguiente medio segundo (señal "chirp UD"), y finalmente la mezcla de un tono puro de 440Hz, una señal de FM con idénticas características frecuenciales a la anterior y un tono de 5kHz, cada una de éstas componentes con una amplitud instantánea distinta a la anterior.

En cuanto a las envolventes propiamente dichas, todas las señales excepto la última presentan envolventes equivalentes: están normalizadas a 0.99, y presentan sendos semienventanados de Hanning al inicio y al final de la señal, de 1/20 de la duración temporal de

[7] *Esta ventana de análisis no es la genérica de Matlab. La genera la propia HRSR.*

[8] *Los puntos por debajo de esta cantidad (en dB) respecto al máximo del escalograma serán silenciados (no se tendrán en cuenta en el análisis).*

Características de las señales			
Señal	**Ley de modulación (fase)**	**Parámetros**	**Duración(s)**
Tono de 440Hz	$2\pi f_1 t$	$f_1 = 440$	0.5
FM $^{(**)}$	$2\pi f_c t + \frac{B}{f_m}\sin(2\pi f_m t)$	$f_c = 1000$, $f_m = 25$, $B = 200$	0.5
LC	$\alpha t^2 + \beta t + \gamma$	$\alpha = 200\pi$, $\beta = 15800\pi$, $\gamma = 0$	0.5
QC	$\alpha t^3 + \beta t^2 + \gamma t + \delta$	$\alpha = \frac{63200\pi}{3}$, $\gamma = 200\pi$, $\beta = \delta = 0$	0.5
HC $^{(*)}$	$2\pi\frac{\alpha}{\beta - t}$	$\beta = 2\pi\frac{(40+\sqrt{20})}{79}$, $\alpha = 200\pi\beta^2$	0.5
Chirp UD	$2\pi\frac{\alpha}{\beta-t}\|_{t=0}^{t=0,5s}$, $\gamma t^2 + \delta t + \zeta\|_{t=0,5}^{t=1s}$	$\alpha, \beta^{(*)}$, $\gamma = -15800\pi$, $\delta = 31800\pi$, $\zeta = 0$	1
Tres tonos	$2\pi f_i t\ \forall i = 1,2,3$	$f_1 = 400$, $f_2 = 600$, $f_3 = 800$	1
FM+440+5kHz	$2\pi f_i(t)$	$f_1(t) = 440t$, $f_2(t)^{(**)}$, $f_3(t) = 5000t$	1

Tabla 5.2: *Descripción de las señales sintéticas analizadas. α, β, γ, δ y ζ se dan en rad/s. Las unidades de f_1, f_2, f_3, f_c, f_m y B son Hz. Los parámetros marcados con $^{(*)}$ y $^{(**)}$ tienen los mismos valores que en las señales con las que se relacionan.*

la misma cada uno. La última señal tiene esta misma envolvente para el tono de 440Hz, una ventana de Hanning equivalente para el tono de FM, pero que entra 2/10 de la duración de la señal más tarde, y el tono de 5kHz enventanado del mismo modo, pero situando la ventana de Hanning entre los 8/10 y los 9/10 de la señal (la duración de este tono es, por lo tanto, muy corta comparada con las demás componentes, por lo que su energía es bastante baja).

Mediante la TFTB de Auger se analiza cada una de estas señales, obteniéndose, en primer lugar, el espectrograma regular STFT, al que se designará por SP, el espectrograma regular reasignado (RSP) y la representación crestas-esqueleto de la señal (RSPr). A continuación se calculan la Distribución Pseudo Wigner-Ville (PWVD), su versión reasignada (RPWVD) y una nueva representación crestas-esqueleto partiendo de estos últimos datos (RPWVDr). De cada una de estas seis representaciones tiempo-frecuencia, se extraen los datos de frecuencia instantánea, comparando los resultados con el valor teórico, perfectamente conocido. Cabe destacar que en las representaciones cresta-esqueleto se tiende a contar con muchos menos puntos que los constituyentes de la señal, en el caso de la RSPr, y más puntos en el caso de la RPWVDr (aunque hay excepciones). Las frecuencias instantáneas se compararán en los instantes temporales donde estén definidas, y en el caso de multi-valuación, se tomará el valor de frecuencia más próximo al dato teórico correspondiente.

Por último, empleando el cálculo del espectro de Nelson de Fulop, se obtiene una vez más un resultado experimental comparable con el teórico allá donde ambos estén definidos.

5.3.3.1. Resultados gráficos

Como primer ejemplo del estudio comparativo efectuado, en las Figuras 5.5(a) a 5.5(c) se muestra la frecuencia instantánea teórica y las experimentales obtenidas a través de la STFT (SP, RSP, RSPr), para el caso de la señal del chirp hiperbólico. En las Figuras 5.6(a) a 5.6(c) se presentan los resultados de la misma señal analizada mediante la PWVD (PWVD, RPWVD y RPWVDr), en la Figura 5.7 los resultados de la HRSR, y por último, en la Figura 5.8 se presenta el resultado comparativo obtenido por mediación de CWAS. Cada una de las gráficas de éstas figuras contiene un pequeño zoom de los resultados experimentales.

Respecto a la Figura 5.5, cabe destacar la poca cantidad de puntos que componen el esqueleto de la señal (marcado con cruces rojas en la Figura 5.5(c)). Este detalle es bastante común en las señales analizadas. La parte de cada gráfica ampliada se corresponde con los instantes de tiempo situados entre los 0.48 y los 0.482 segundos. En estos instantes de tiempo, la pendiente de la ley frecuencial (hipérbola) ya es muy elevada, con lo cual los errores tienden a ser grandes.

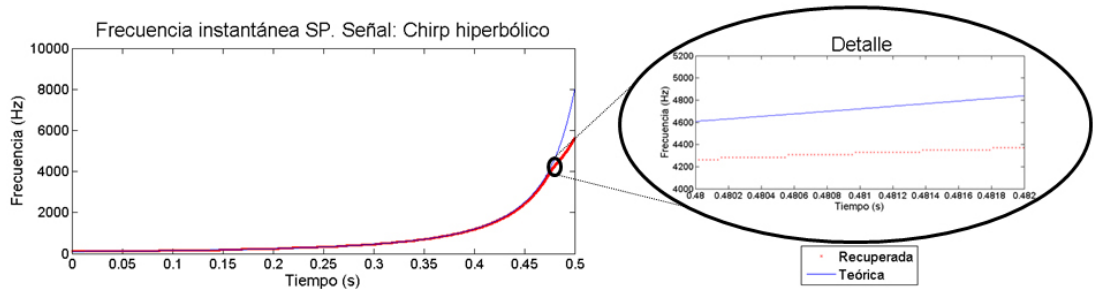
En cuanto a los resultados mostrados en la Figura 5.6, el esqueleto original de la señal presenta un elevado número de instantes temporales multivaluados. En el esqueleto final, representado con cruces rojas en la Figura 5.6(c), se ha tomado el valor de la frecuencia más cercano al valor teórico para los casos de éstas frecuencias múltiples, como se ha explicado anteriormente. La parte de cada gráfica ampliada se corresponde de nuevo con los instantes de tiempo situados entre los 0.48 y los 0.482 segundos, con lo que el error cometido tiende a ser más elevado.

En la Figura 5.7 se presentan los resultados correspondientes a la recuperación de la frecuencia instantánea empleando la rutina de Fulop (HRSR) para la misma señal. Obsérvese que el conjunto final de puntos experimentales es tan pequeño que la sección detallada en la parte derecha de la figura varía entre los 0.4 y los 0.49 segundos, de cara a presentar un número suficiente de puntos como para resultar representativo.

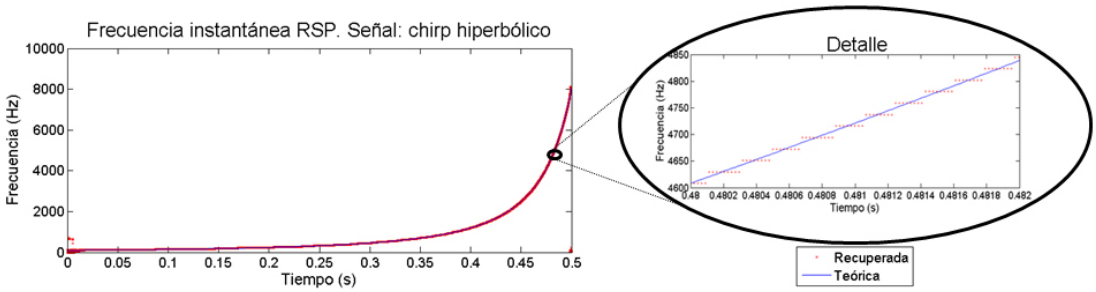
Por último, en los resultados mostrados en la Figura 5.8 se puede apreciar cómo, pese a que los instantes de tiempo ampliados se corresponden con los de las Figuras 5.5 y 5.6 (de 0.48 a 0.482 segundos), los errores continúan siendo demasiado pequeños para resultar visibles con este nivel de detalle. Esto es una prueba parcial que demuestra la mayor precisión del algoritmo CWAS.

Comparando las Figuras 5.5 a 5.8, se puede deducir que la técnica más cercana en precisión al algoritmo CWAS es el algoritmo de Fulop, si bien sería necesario interpolar los valores de frecuencia para obtener un número de puntos significativo. Estas conclusiones se verán apoyadas por las gráficas siguientes y sobre todo los resultados numéricos.

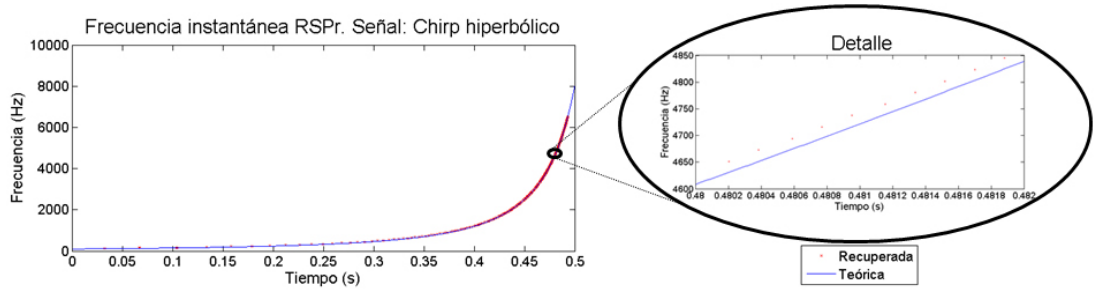
En las Figuras 5.9 a 5.12 se presentan algunos de los resultados obtenidos para el caso de la señal mezcla de FM, tono de 440Hz y tono de 5kHz. Éstos resultados se limitan a los



(a) f_{ins}: *SP('HC').*



(b) f_{ins}: *RSP('HC').*

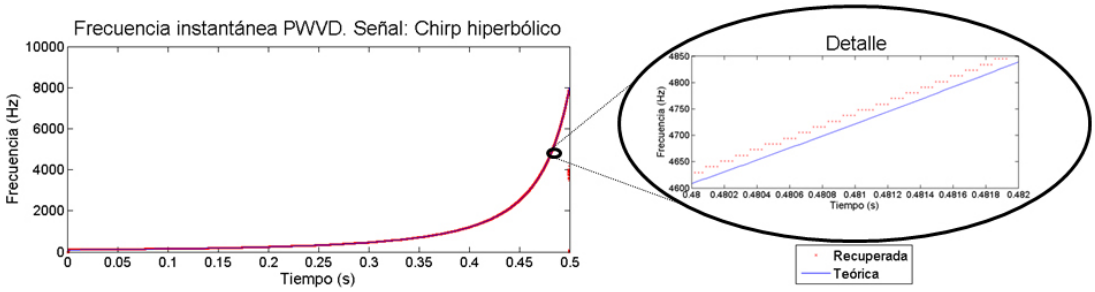


(c) f_{ins}: *RSPr('HC').*

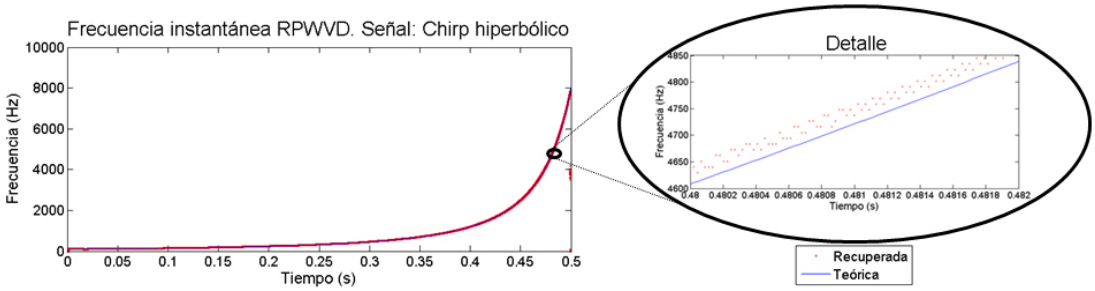
Figura 5.5: *Comparativas entre el valor teórico de la frecuencia instantánea (trazo azul continuo) y los valores recuperados a través de los diferentes espectrogramas STFT (cruces rojas). Señal: chirp hiperbólico. (a) Espectrograma regular. (b) Espectrograma STFT reasignado. (c) Esqueleto del espectrograma STFT reasignado.*

esqueletos de la señal (RSPr y RPWVDr) como representación de la TFTB, a los resultados por el algoritmo de Fulop, y a los resultados obtenidos empleando CWAS.

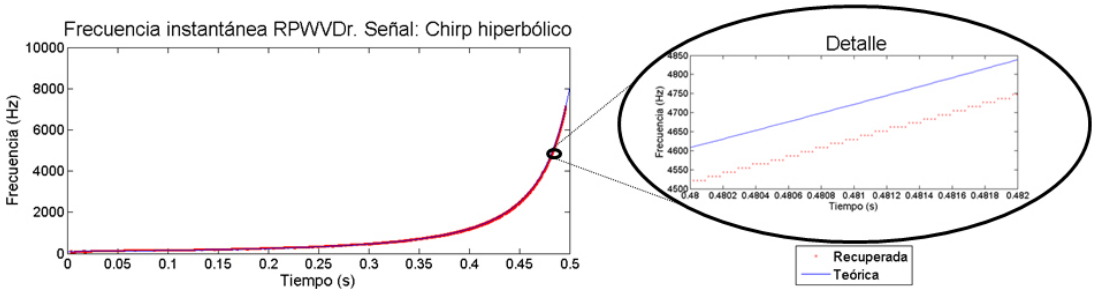
Aunque las diferentes frecuencias instantáneas teóricas (en azul) se han representado para la completa longitud de la señal, en la figuras correspondientes a los resultados de la PWVD, las frecuencias instantáneas experimentales correspondientes a las componentes de FM y el tono de 5kHz han sido adecuadas a la región de existencia de sus correspondientes



(a) f_{ins}: *PWVD('HC').*



(b) f_{ins}: *RPWVD('HC').*



(c) f_{ins}: *RPWVDr('HC').*

Figura 5.6: *Comparativas entre el valor teórico de la frecuencia instantánea (trazo azul continuo) y los valores recuperados a través de la PWVD (cruces rojas). Señal: chirp hiperbólico. (a) Distribución Pseudo Wigner-Ville. (b) PWVD reasignada. (c) Esqueleto de la PWVD reasignada.*

amplitudes. Estas regiones de existencia aparecen delimitadas por líneas negras verticales discontinuas en las Figuras 5.9(c), 5.9(d), 5.10(c), 5.10(d), 5.12(c) y 5.12(d).

Obsérvese, en la Figura 5.9(d), la muy limitada región del semiplano donde el esqueleto del espectrograma regular reasignado ofrece resultados (círculo rojo discontinuo).

En la Figura 5.10(a) se han marcado, de nuevo con cruces rojas, los valores experimentales de las frecuencias instantáneas correspondientes a las tres componentes de la señal. Con

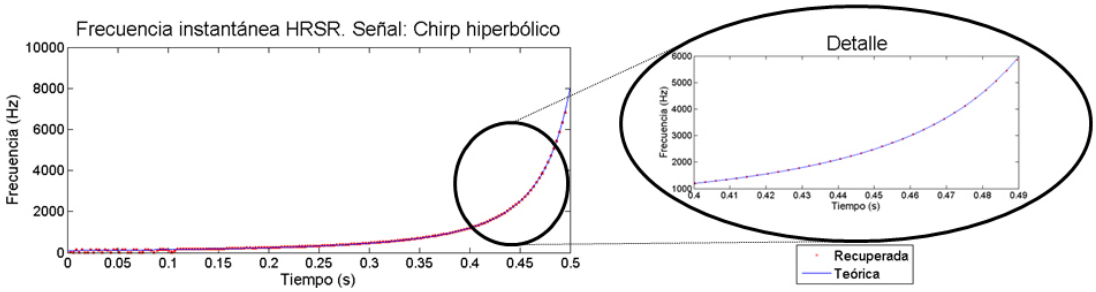


Figura 5.7: *Comparativa entre el valor teórico de la frecuencia instantánea (trazo azul continuo) y el valor recuperado por el algoritmo de Fulop (cruces rojas). Señal: chirp hiperbólico.*

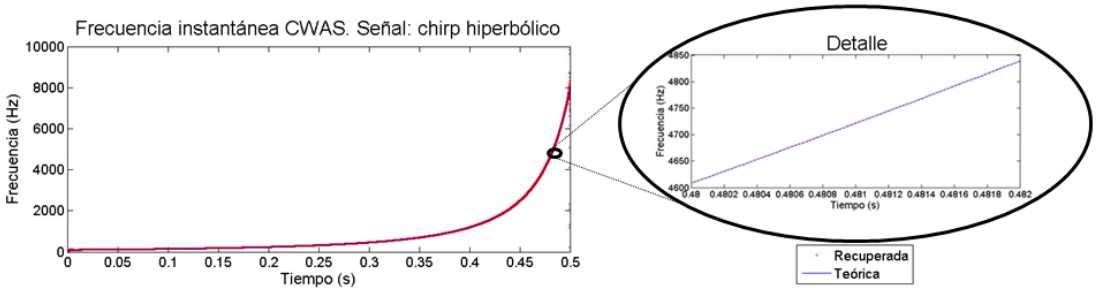


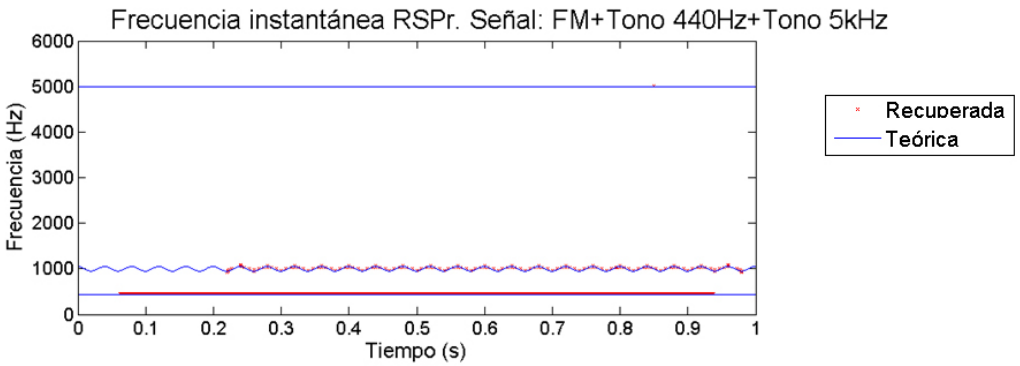
Figura 5.8: *Comparativa entre el valor teórico de la frecuencia instantánea (trazo azul continuo) y el valor recuperado a través de CWAS (cruces rojas). Señal: chirp hiperbólico.*

círculos magenta aparecen los términos de interferencia. Estos términos se corresponden con cada una de las posibles parejas de componentes que se tomen, y aparecen aproximadamente en la zona de frecuencias intermedias de los términos interferentes y para los tiempos en que ambas componentes coexisten. Como se ha dicho, esto dificulta la visualización e interpretación de la información proporcionada por la PWVD.

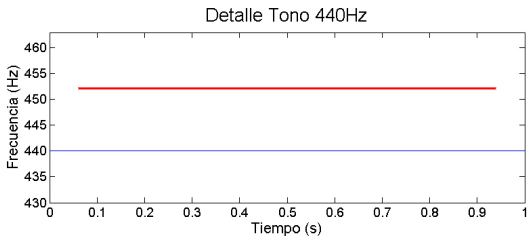
En cuanto a los resultados obtenidos mediante las rutinas de alta resolución de Fulop, aparecen en la Figura 5.11.

Para terminar, el algoritmo CWAS ofrece las frecuencias instantáneas de sus tres parciales más energéticos que se muestran en la Figura 5.12(a). Como siempre, en las zonas donde la amplitud está por debajo de cierto valor, la frecuencia instantánea empieza a no estar bien definida, apareciendo oscilaciones.

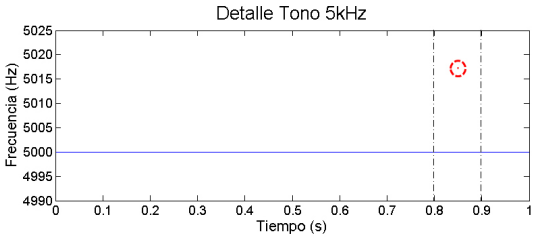
Tales oscilaciones se hacen más evidentes en las Figuras 5.12(b) y 5.12(c). Sin embargo,



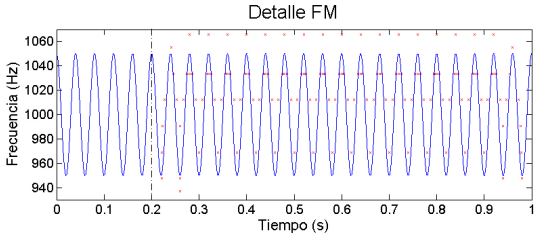
(a) f_{ins}: *RSPr. Señal: FM + Tono 440Hz + Tono 5kHz.*



(b) f_{ins}: *RSPr(Tono 440Hz).*



(c) f_{ins}: *RSPr(5kHz).*

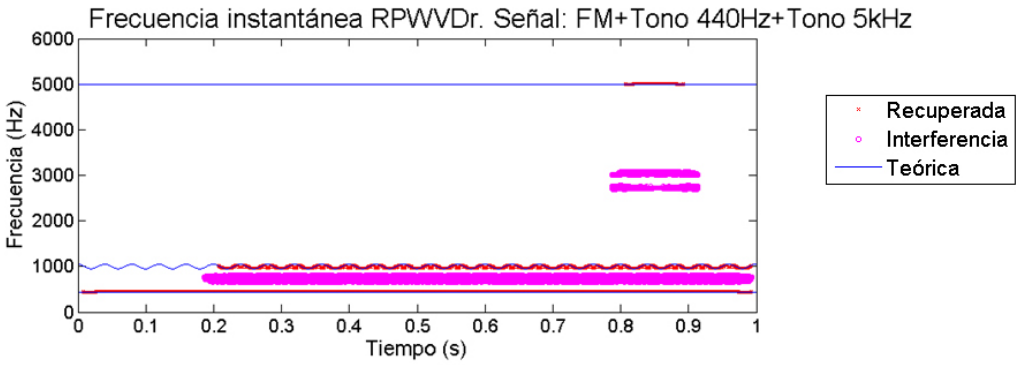


(d) f_{ins}: *RSPr(FM).*

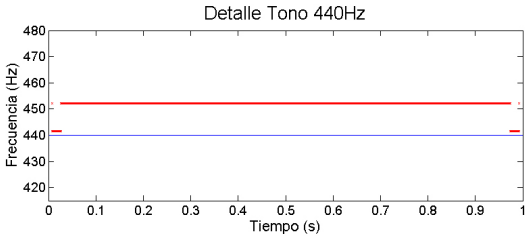
Figura 5.9: *Señal: FM + Tono 440Hz + Tono 5kHz. (a) Comparativa entre los valores teóricos de la frecuencia instantánea (trazos azules) y los recuperados a través del esqueleto del espectrograma reasignado (cruces rojas). (b) a (d) Detalles de la comparativa a través de la RSPr. (b) Tono de 440Hz. (c) Tono de 5kHz (resultados experimentales enmarcados por un círculo rojo discontinuo). (d) Componente de FM.*

la amplitud de éstas es, en este caso, muy baja (obsérvese el valor del eje vertical en cada subfigura). La precisión en la captura de la frecuencia instantánea puede verse especialmente en el zoom de la componente de FM, mostrado en la Figura 5.12(d).

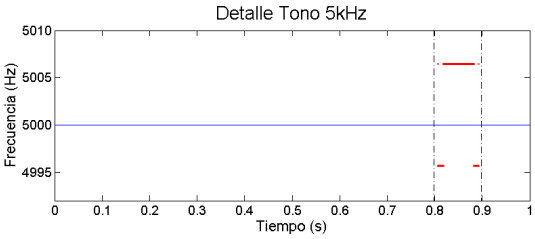
Las dos señales representadas (chirp hiperbólico y mezcla de FM, tono de 440Hz y tono de 5kHz) son meramente representativas. Las figuras mostradas en este apartado se complementan con los resultados numéricos del siguiente, así como con los espectrogramas



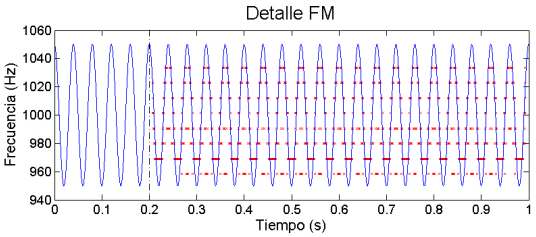
(a) f_{ins}: *RPWVDr. Señal: FM + Tono 440Hz + Tono 5kHz.*



(b) f_{ins}: *RPWVDr(Tono 440Hz).*



(c) f_{ins}: *RPWVDr(5kHz).*



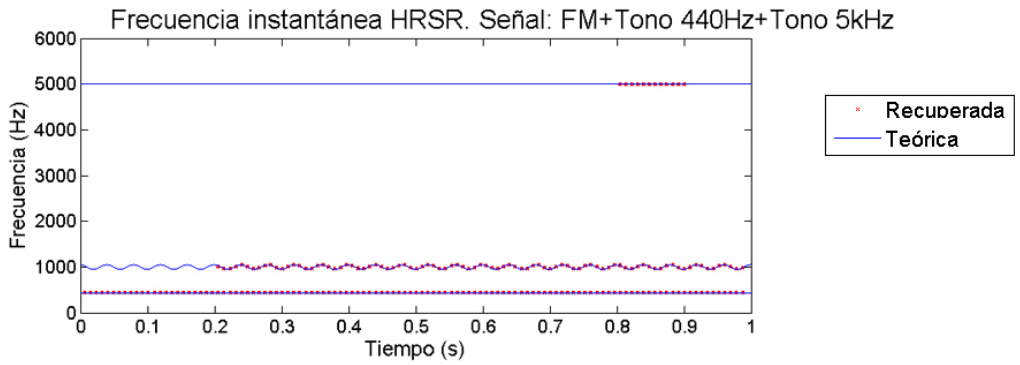
(d) f_{ins}: *RPWVDr(FM).*

Figura 5.10: *Señal: FM + Tono 440Hz + Tono 5kHz. (a) Comparativa entre el valor teórico de la frecuencia instantánea (trazos azules) y el valor recuperado a través del esqueleto de la PWVD reasignada (cruces rojas). Los términos de interferencia aparecen marcados con círculos magenta. (b) a (d) Detalles de los valores experimentales de la RPWVDr. (b) Tono de 440Hz. (c) Tono de 5kHz. (d) Componente de FM.*

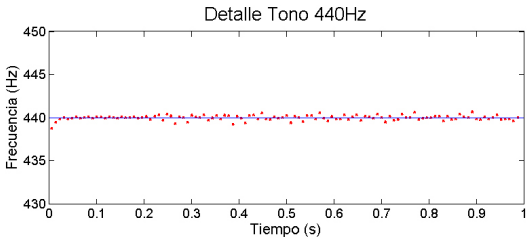
en dos y tres dimensiones que se presentarán en la Sección 5.4, donde se mostrarán además resultados relativos a señales de audio no sintéticas.

5.3.3.2. Valores numéricos

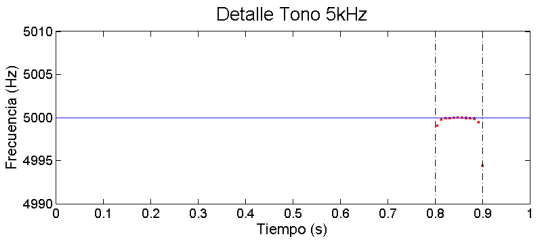
De cara a obtener un resultado numérico correspondiente a la magnitud del error cometido en la obtención de la frecuencia instantánea, en primer lugar se evaluará para cada



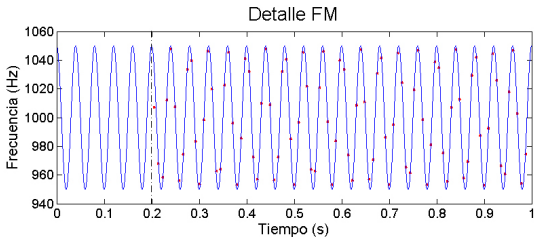
(a) f_{ins}: *HRSR(Señal: FM + Tono 440Hz + Tono 5kHz.).*



(b) f_{ins}: *HRSR(Tono 440Hz).*



(c) f_{ins}: *HRSR(5kHz).*

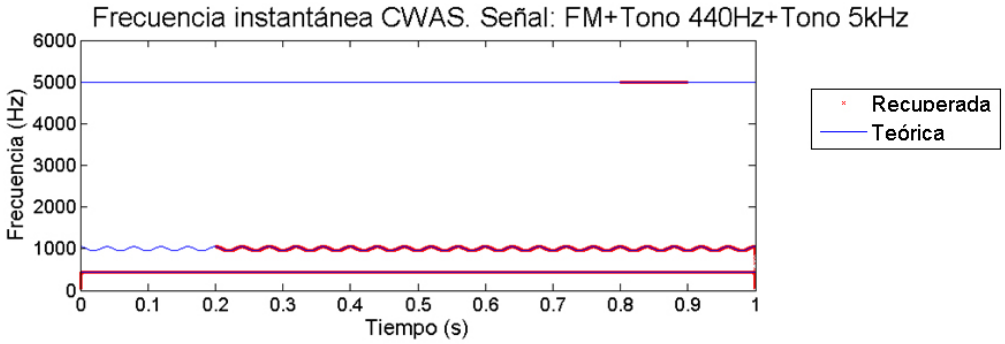


(d) f_{ins}: *HRSR(FM).*

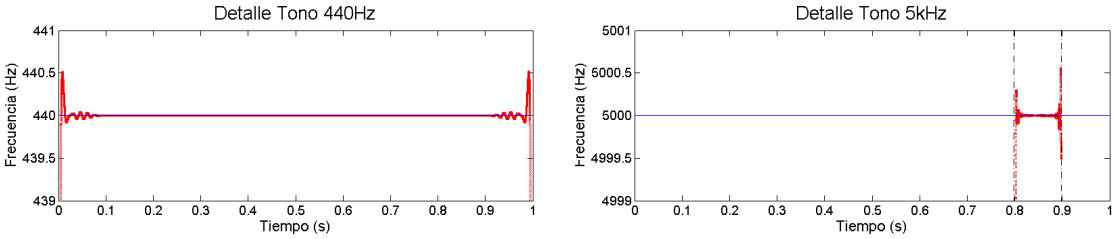
Figura 5.11: *Señal: FM + Tono 440Hz + Tono 5kHz. (a) Comparativa entre el valor teórico de la frecuencia instantánea (trazos azules) y el valor recuperado a través del espectrograma de alta resolución de Fulop (cruces rojas). (b) a (d) Detalles de los valores experimentales recuperados a través del TCIFS. (b) Tono de 440Hz. (c) Tono de 5kHz. (d) Componente de FM.*

método la diferencia entre los datos experimentales y los correspondientes datos teóricos. El dato de error característico será el promedio de estos datos puntuales.

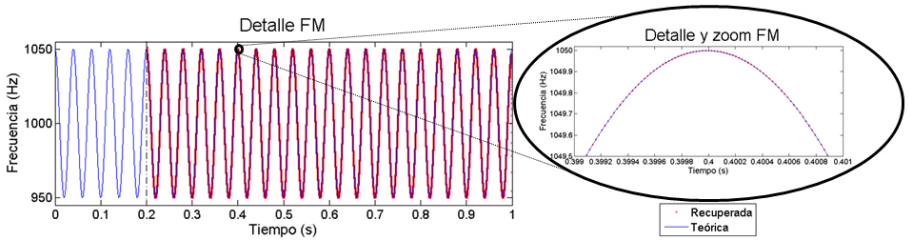
Más en concreto, para cada señal $x(t)$ y para cada distribución tiempo-frecuencia, los



(a) f_{ins}: *CWAS(Señal: FM + Tono 440Hz + Tono 5kHz).*



(b) f_{ins}: *CWAS(Tono 440Hz).* (c) f_{ins}: *CWAS(5kHz).*



(d) f_{ins}: *CWAS(FM).*

Figura 5.12: *Señal: FM + Tono 440Hz + Tono 5kHz. (a) Comparativa entre el valor teórico de la frecuencia instantánea (trazos azules) y el valor recuperado a través de CWAS (cruces rojas). (b) a (d) Detalles de la comparativa de resultados a través del algoritmo CWAS. (b) Tono de 440Hz. (c) Tono de 5kHz. (d) Componente de FM.*

errores en frecuencia se calculan en tanto por ciento (%), según la expresión:

$$\epsilon_{f_{ins}}^{(TFDx)}(t_k) = 100 * \frac{|f_{teo}^{(x)}(t_k) - f_{ins}^{(TFDx)}(t_k)|}{f_{teo}^{(x)}(t_k)} \tag{5.5}$$

Esta ecuación se calcula únicamente en el conjunto de puntos $\{t_k\}$ donde esté definida la frecuencia instantánea experimental de la señal $x(t)$ a través de la transformada TFD, $f_{ins}^{(TFDx)}$. Como se ha avanzado anteriormente, este conjunto de puntos puede ser sensible-

mente inferior a la longitud real de la señal, o puede darse el caso de instantes temporales multivaluados (en cuyo caso se toma el dato más aproximado al valor teórico de la señal en ese punto).

En cuanto al valor medio del error, puede escribirse como:

$$\overline{\epsilon_{f_{ins}}^{(TFDx)}}(t_k) = mean\Big|_{t_k}\left[\epsilon_{f_{ins}}^{(TFDx)}(t_k)\right] \tag{5.6}$$

En la Tabla 5.3 se recogen los resultados experimentales de error en la frecuencia instantánea obtenidos a través de las Ecuaciones (5.5) y (5.6), para las ocho señales sintéticas analizadas mediante cada una de las siete herramientas detalladas anteriormente (es decir, SP, RSP, RSPr, PWVD, RPWVD, RPWVDr y HRSR). La última columna de la tabla refleja los resultados de error medio en la obtención de la frecuencia instantánea obtenidos a través del algoritmo CWAS.

	Errores experimentales (%)							
Señal	**SP**	**RSP**	**RSPr**	**PWVD**	**RPWVD**	**RPWVDr**	**HRSR**	**CWAS**
Tono 440	2.12	3.90	2.77	0.34	0.40	2.77	6.91E-2	1.48E-2
FM	1.51	1.49	1.56	1.08	1.06	1.16	5.38E-2	6.7E-3
LC	2.60	1.05	0.54	0.26	0.28	0.27	6.34E-2	4.9E-3
QC	1.05	2.50	0.55	0.62	0.70	0.32	0.26	9.9E-3
HC	3.08	3.27	3.27	0.73	1.16	3.27	1.01	1.72E-2
Chirp UD	3.06	2.49	0.78	0.71	0.82	0.89	0.47	3.06E-2
Tres tonos:								
Tono 400Hz	2.33	2.48	7.67	0.44	1.02	2.28	0.70	4.93E-6
Tono 600Hz	2.66	2.79	4.08	2.04	1.34	2.28	0.70	3.29E-7
Tono 800Hz	2.25	1.88	2.28	1.52	1.14	0.94	1.37	3.41E-5
FM+440+5k:								
Tono 440Hz	2.12	2.12	2.77	0.33	0.33	2.77	4.75E-2	1.49E-5
FM	1.41	1.37	7.86	1.15	1.15	3.50	0.20	2.68E-4
Tono 5kHz	0.09	0.09	0.34	0.09	0.09	0.13	4.2E-3	6.77E-5

Tabla 5.3: *Comparativa de los errores en la detección de frecuencia instantánea cometidos mediante las 8 rutinas empleadas. Los datos aparecen en tanto por ciento (%).*

A la luz de los datos arrojados por esta tabla, el algoritmo CWAS resulta en el peor de los casos un orden de magnitud más preciso que cualquiera de las técnicas en la recuperación de la frecuencia instantánea de la señal basadas en la TFTB de Auger. La misma conclusión se extrae de la comparativa con las HRSR de Fulop (excepto el tomo de 440Hz, dónde los resultados de CWAS son *tan sólo* medio orden de magnitud mejores). Las frecuencias instantáneas de las señales de complejidad creciente (la señal de los tres tonos puros y la

mezcla de FM, tono de 440Hz y tono de 5kHz) son mejor recuperadas por el algoritmo CWAS pese a que los resultados del algoritmo de Fulop son también muy precisos.

5.4. Representación tiempo–frecuencia: Visualización

Antes de abordar la comparativa en la resíntesis de la señal de audio, conviene presentar algunos resultados experimentales que refuerzan la solidez del método presentado. Estos resultados abarcan la capacidad de representación de la información proporcionada por el algoritmo. En este caso, se va a comparar la representación tiempo–frecuencia propuesta en esta Tesis (módulo y frecuencia instantáneos de los parciales) frente al espectrograma de alta resolución de Fulop, tanto en dos como tres dimensiones. En un paso previo, se mostrarán espectrogramas wavelet obtenidos mediante el algoritmo CWAS en comparación con los espectrogramas regular y procedente de la PWVD, también en dos y en tres dimensiones. El grupo de señales analizadas comprende tanto las señales sintéticas de la Sección 5.3.3 como grabaciones de instrumentos reales.

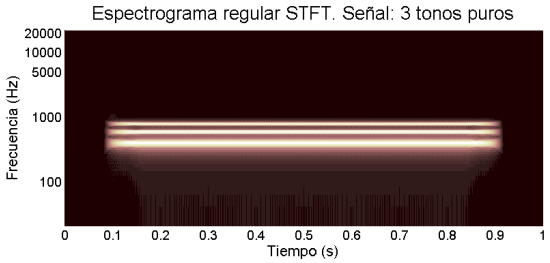
5.4.1. Espectrogramas

En efecto, la comparativa entre los espectrogramas regular FFT, de la PWVD y wavelet resulta en sí misma significativa. Se van a presentar tres escalogramas representativos de las señales analizadas, tanto en dos como en tres dimensiones. El objetivo es demostrar que, pese a que el algoritmo CWAS presenta una quinta parte de los bins frecuenciales que la STFT y la PWVD representadas, la información gráfica wavelet es de interpretación, precisión y calidad definitivamente superiores a la PWVD, y cuando menos comparables a la STFT. Las tres señales escogidas para tal efecto son una señal sintética (en concreto la mezcla de tres sinusoides de 400, 600 y 800Hz) y dos grabaciones reales de audio (una guitarra $B4$ con un bending muy marcado y un saxo tocando una $C3$).

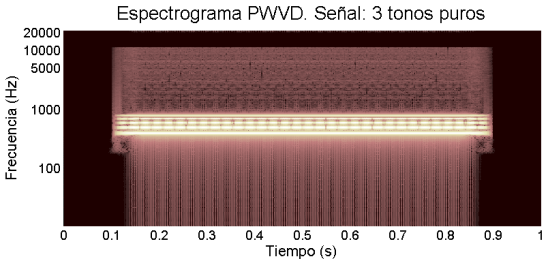
5.4.1.1. Representación plana

A lo largo de esta disertación, se han mostrado diversos espectrogramas wavelet en dos dimensiones. En ellos se aprecian las trayectorias claras correspondientes a los parciales de cada señal analizada. Asimismo, ya dentro del presente Capítulo, en las Secciones 5.2.1 y 5.2.2, se han mostrado sendos espectrogramas planos (regular y wavelet) correspondientes a la misma señal. A continuación se mostrarán los espectrogramas bidimensionales de las tres señales antes mencionadas (en dB), obtenidos mediante la STFT puntual (programada por el autor, Sección 5.2.1), la PWVD de Auger y el algoritmo CWAS.

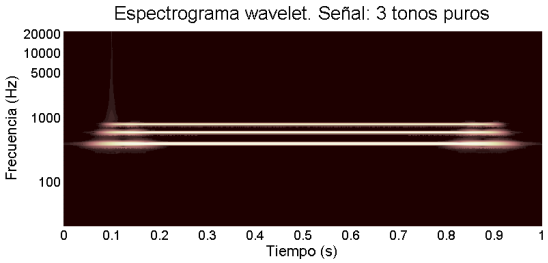
En cuanto a la señal sintética, los resultados se muestran en la Figura 5.13.



(a) *Espectrograma plano regular.*



(b) *Espectrograma plano PWVD.*



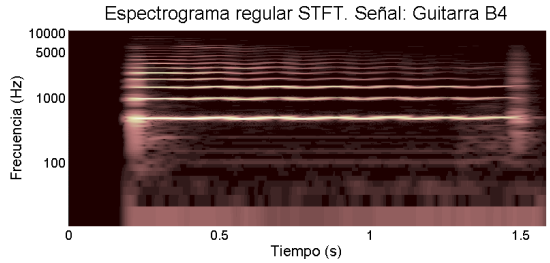
(c) *Espectrograma wavelet plano.*

Figura 5.13: *Espectrogramas bidimensionales. Señal: tres tonos puros de frecuencias 400Hz, 600Hz y 800Hz. (a) Espectrograma regular plano basado en la STFT. (b) Espectrograma plano basado en la PWVD. (c) Espectrograma plano CWAS.*

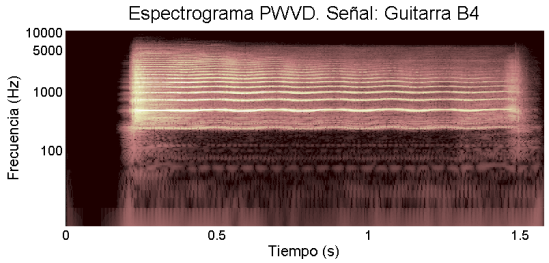
En esta figura, el espectrograma wavelet ha sido obtenido bajo las condiciones normales de análisis (en cuanto al número de divisiones por octava D para una frecuencia de muestreo $f_s = 44100$Hz), lo que arroja un total de 201 bandas de frecuencia. Con 2048 bins frecuenciales en la STFT y la PWVD, se obtienen 1048 bandas de análisis. Como se puede observar, el espectrograma wavelet localiza los tres tonos con más limpieza que la STFT, marca mejor los transitorios y no presenta tanto ruido de baja frecuencia. Respecto a la PWVD, las ventajas son aún más evidentes. La PWVD localiza cinco tonos (dos de ellos se corresponden con las interferencias cruzadas) y el fondo resulta sensiblemente más ruidoso en prácticamente todo el plano.

Algunas de estas ventajas deberían resultar más evidentes para señales más complicadas, como por ejemplo señales de audio reales. En las Figuras 5.14 y 5.15 se muestran los espectrogramas bidimensionales de la guitarra $B4$ y del saxo $C3$.

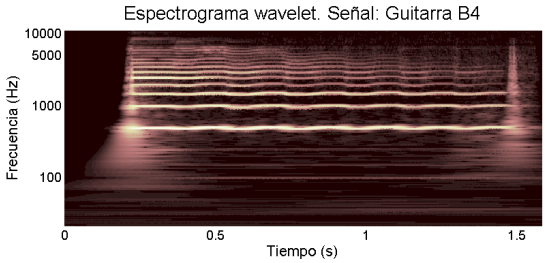
De éstas figuras se desprenden varias conclusiones interesantes. En primer lugar, comparando las Figuras 5.14 y 5.15 (a) y (c), la fundamental de cada señal, así como sus primeros armónicos, son al menos tan evidentes en el análisis wavelet propuesto como en el STFT clásico. El algoritmo CWAS localiza tanto los transitorios como las componentes sinusoidales de la señal bajo las mismas condiciones de análisis. La STFT necesitaría ventanas de dife-



(a) *Espectrograma plano regular.*



(b) *Espectrograma plano PWVD.*



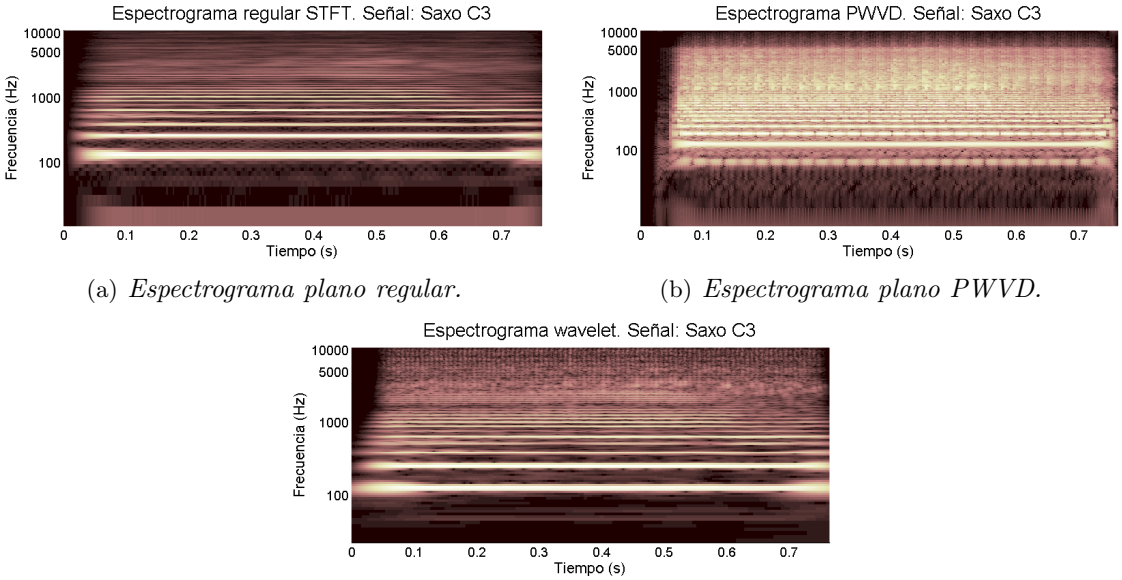
(c) *Espectrograma wavelet plano.*

Figura 5.14: *Espectrogramas bidimensionales. Señal: Guitarra B4. (a) Espectrograma regular plano basado en la STFT. (b) Espectrograma plano basado en la PWVD. (c) Espectrograma plano CWAS.*

rentes tamaños para localizar suficientemente bien los eventos temporales, como los ataques de las notas, y las componentes frecuenciales involucradas en éstos. En éstas figuras, el eje frecuencial logarítmico no permite realizar fácilmente una comparativa de la zona alta del espectro, lo que queda pendiente para la Sección 5.4.1.2. Sin embargo, por el mismo motivo, puede verse fácilmente que tanto la STFT como la PWVD presentan una menor resolución en la parte baja del espectro. Puede verse cómo el algoritmo CWAS localiza de forma más definida las componentes (ruidosas o subarmónicas) por debajo de la fundamental.

Comparando las Figuras 5.14 y 5.15 (b) y (c), se aprecia cómo el espectrograma wavelet es sensiblemente más limpio que el basado en la PWVD. Entre las verdaderas componentes sinusoidales de la señal pueden apreciarse un gran número de espurios (interferencias) que generan confusión en el espectrograma. De este modo, localizar y extraer la información tempo-frecuencial de la señal puede resultar bastante complicado, como se detalla en los trabajos de Boashash [31, 32].

En el Anexo IV.b se presentan más resultados relacionados con el conjunto de señales analizadas (en concreto una comparativa de los escalogramas en dos y tres dimensiones de cinco de éstas, obtenidos mediante las tres técnicas).



(a) *Espectrograma plano regular.*

(b) *Espectrograma plano PWVD.*

(c) *Espectrograma wavelet plano.*

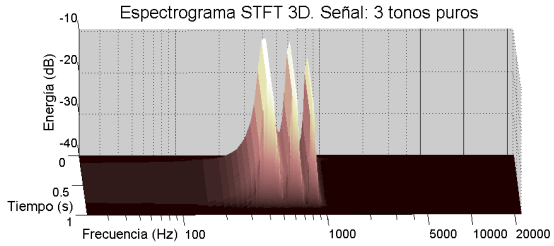
Figura 5.15: *Espectrogramas bidimensionales. Señal: Saxo C3. (a) Espectrograma regular plano basado en la STFT. (b) Espectrograma plano basado en la PWVD. (c) Espectrograma plano CWAS.*

5.4.1.2. Representación volumétrica

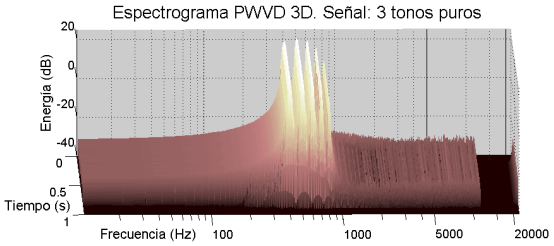
Levantando en el eje Z el valor del módulo de los coeficientes de las transformadas (el valor de los coeficientes en sí para la PWVD, ya que ésta es definida real) se obtiene una superficie tridimensional que refleja el valor en amplitud del comportamiento frecuencial de cada componente caracterizada.

En la Figura 5.16 aparecen las representaciones tridimensionales correspondientes a la señal mezcla de tres tonos puros.

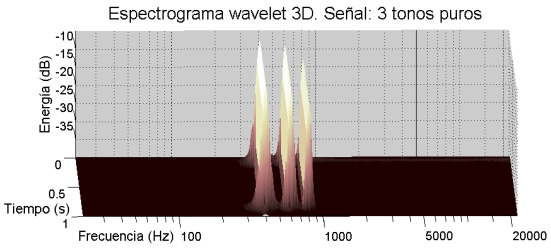
Como se puede observar, el algoritmo CWAS proporciona un escalograma más definido que el obtenido a través de la STFT en este caso, puesto que las frecuencias de las componentes de la señal son relativamente bajas. Con respecto al escalograma basado en la PWVD, pueden observarse los cinco picos anteriormente mencionados. Comparando las alturas de los tres picos principales en los tres espectrogramas, resulta evidente que, para el caso de la PWVD, el pico de 600Hz se va reforzado por la interferencia de los términos de 400Hz y 800Hz, lo que dificultaría enormemente la tarea de obtener con precisión la envolvente instantánea de la señal para este tono. Puede apreciarse además el fondo ruidoso, mucho más presente en la Figura 5.13(b) que en las demás.



(a) *Espectrograma regular 3D.*



(b) *Espectrograma PWVD 3D.*



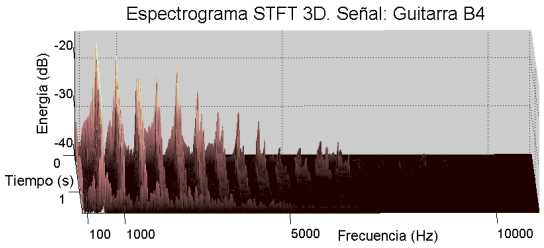
(c) *Espectrograma wavelet 3D.*

Figura 5.16: *Espectrogramas tridimensionales. Señal: tres tonos puros de frecuencias 400Hz, 600Hz y 800Hz. (a) Espectrograma regular 3D basado en la STFT. (b) Espectrograma 3D basado en la PWVD. (c) Espectrograma CWAS 3D.*

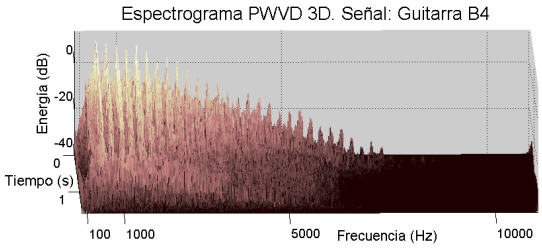
En las Figuras 5.17 y 5.18 se han representado los espectrogramas 3D correspondientes a la guitarra y el saxo.

Comparando las Figuras 5.17 y 5.18, (a) y (c), se puede ver cómo en la parte baja-media del espectro (2 ó 3kHz), la STFT y el algoritmo CWAS ofrecen resultados similares en cuanto al número y cualidades principales de las componentes encontradas. Sin embargo, en la parte alta del espectro, la STFT presenta una mayor resolución. Esto es debido al hecho ya comentado del equiespaciado del eje frecuencial en los algoritmos FFT, mientras que CWAS presenta un eje frecuencial de resolución logarítmica (Q constante). Las diferencias se hacen más notables en la Figura 5.18 correspondiente a la señal del saxo (ya que el número de armónicos es bastante más elevado, dada la menor fundamental de la nota ejecutada). Esta falta de resolución en la técnica presentada es relativamente fácil de evitar (bajo coste en tiempo de computación), y en cualquier caso no representa una dificultad seria para ninguna de las aplicaciones desarrolladas.

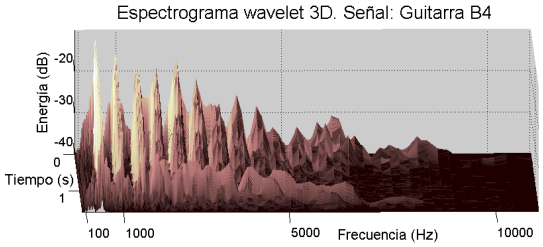
En las Figuras 5.17 y 5.18, (b) y (c), que representan los escalogramas 3D wavelet y PWVD, las ventajas de la técnica propuesta se hacen más evidentes. Puede observarse con facilidad que el número de componentes detectadas con la PWVD es mucho mayor de lo esperado, debido a los términos de interferencia, que en señales reales pueden llegar a suponer un verdadero problema, difícil de resolver.



(a) *Espectrograma regular 3D.*



(b) *Espectrograma PWVD 3D.*



(c) *Espectrograma wavelet 3D.*

Figura 5.17: *Espectrogramas tridimensionales. Señal: Guitarra B4. (a) Espectrograma regular 3D basado en la STFT. (b) Espectrograma 3D basado en la PWVD. (c) Espectrograma CWAS 3D.*

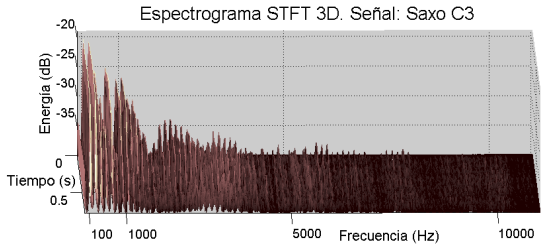
5.4.2. Modelo de la señal

Como se ha documentado ampliamente, el algoritmo CWAS despliega una elevada capacidad para encontrar, seguir y caracterizar dinámicamente las diferentes componentes frecuenciales presentes en la señal analizada (adecuadamente separadas por el banco de filtros complejos empleado), ofreciendo una función compleja para cada parcial detectado de la señal. De ésta función compleja, definida a través de la Ecuación (3.18), se pueden obtener la amplitud y fase instantáneas del parcial, a través de las Ecuaciones (3.20) y (3.21) respectivamente. Aplicando la Ecuación (3.23) a la fase desenrollada, se obtiene la frecuencia instantánea.

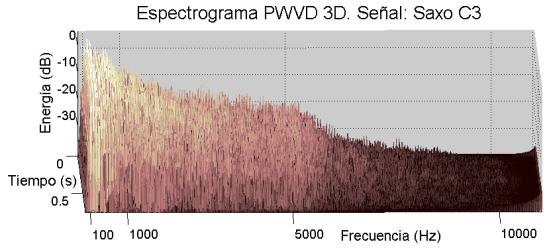
Con ésta información disponible, se hace posible representar gráficamente, en el semiplano Tiempo–Frecuencia y para cada parcial, la terna:

$$\Big(t_i, f_{ins,n}(t_i), A_n(t_i)\Big) \qquad \forall t_i, \quad \forall n \tag{5.7}$$

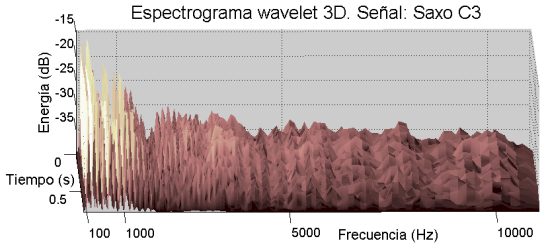
Esto proporciona, de forma directa, una información *equivalente* a la de la frecuencia instantánea canalizada, con la excepción de que representa *por defecto* la información en todos los puntos de existencia de cada uno de los parciales detectados. De la representación



(a) *Espectrograma regular 3D.*



(b) *Espectrograma PWVD 3D.*



(c) *Espectrograma wavelet 3D.*

Figura 5.18: *Espectrogramas tridimensionales. Señal: Saxo C3. (a) Espectrograma regular 3D basado en la STFT. (b) Espectrograma 3D basado en la PWVD. (c) Espectrograma CWAS 3D.*

espectrográfica dispersa por el ancho de banda instantáneo, se pasa a una representación en parciales muy definida, que además retiene en sí misma todas las características de la señal original. En las siguientes Secciones se presenta ésta información en dos y tres dimensiones, comparándola, donde es posible, con la frecuencia canalizada de Fulop.

5.4.2.1. Representación 2D

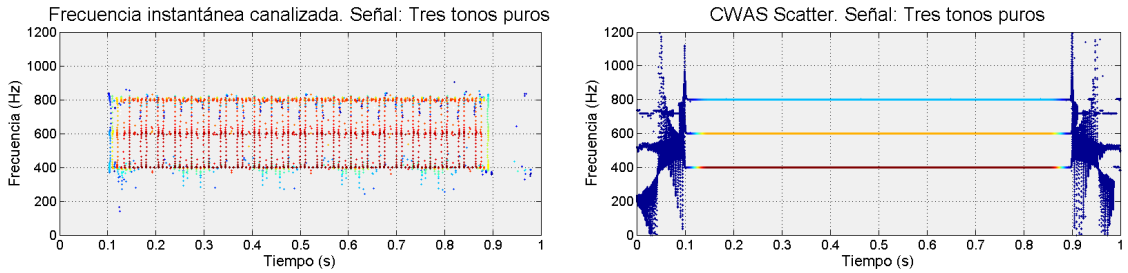
Un primer modo de visualización de esta información consiste en la representación bidimensional de los puntos:

$$\left(t_i, f_{ins,n}(t_i)\right) \tag{5.8}$$

para cada uno de los parciales detectados. En ésta representación bidimensional simple se incluye la amplitud, $A_n(t_i)$, mediante un mapeo de color proporcional al valor de las amplitudes instantáneas de los parciales.

En la Figura 5.19(a) se presentan la información tempo-frecuencial extraída del algoritmo de Fulop para la señal de los tres tonos de 400, 600 y 800Hz. Éstas gráficas se han obtenido mediante la función *scatter3* de Matlab®. Como se puede comprobar fácilmente, aunque la información frecuencial es bastante precisa (véase la Tabla 5.3), la representación visual

directa de la misma no lo es tanto. Ésta es una característica inherente a los espectrogramas reasignados: a pesar de los logros evidentes en la claridad, pueden ser bastante ruidosos, a causa de la aparición de puntos repartidos al azar en las regiones donde los datos reasignados no están claramente asociados con una componente sinusoidal [58].



(a) *CIF de Fulop.* (b) *CWAS TFR2D.*

Figura 5.19: *Representaciones tiempo–frecuencia bidimensionales. Señal: tres tonos puros. (a) Frecuencia instantánea canalizada de Fulop. (b) Representación tiempo–frecuencia CWAS (2D).*

La información en parciales del algoritmo CWAS, Figura 5.19(b), es intrínsecamente más ordenada que la representación directa de los datos que arrojan las HRSR de Fulop, como se deduce comparando las subfiguras. Esto es debido a que los datos representados por el algoritmo CWAS están adecuadamente refinados, mientras que la CIF de Fulop se presenta en bruto. Se podría obtener una representación equivalente (y muy similar en el plano visual) de la frecuencia instantánea canalizada, dividiendo adecuadamente la información en parciales, e interpolando los datos de frecuencia y módulo para adecuarlos a la duración de la señal. Los resultados de las señales del saxo y la guitarra de la siguiente Sección se representarán siguiendo esta técnica.

5.4.2.1.1. Mejoras en la representación visual

La representación directa de la información que arrojan los coeficientes wavelet no resulta especialmente esclarecedora *per se* en las zonas donde la envolvente $A_n(t_i)$ de un parcial determinado sea relativamente baja, (y donde, como se ha repetido en varias ocasiones, la Ecuación (3.23), de la que se extrae su frecuencia instantánea no está bien definida, y por lo tanto ésta presenta muy evidentes oscilaciones, que no son causa final de desvíos en los resultados prácticos, precisamente porque la amplitud de la envolvente asociada en esos puntos es tan baja).

Para evitar esta variabilidad en la representación visual de la información, se puede

recurrir a una técnica de suavizado que incluye un filtrado paso bajo y un subsampleado de muestras con una posterior interpolación. La mayoría de la oscilación presente en los diferentes parciales queda sensiblemente reducida de este modo, lo cual permite mejorar la representación gráfica de la información. Ésta técnica se puede aplicar tanto al algoritmo CWAS como a las HRSR.

En las Figuras 5.20 y 5.21 se han representado gráficamente los resultados comparativos de los algoritmos CWAS y HRSR (CIF interpolada) para el caso de las señales de la guitarra ejecutando una nota $B4$ y el saxo tocando una nota $C3$. En ambas figuras queda patente la mayor resolución de la STFT en alta frecuencia (lo cual no es una ventaja en el campo visual y tampoco se traduce necesariamente en mejores resultados de síntesis, como se demostrará en la Sección 5.5). En la zona media-baja del espectro, ambas representaciones resultan similares, Figuras 5.20(c)-5.20(d) y 5.21(c)-5.21(d). Como se puede observar, la CIF de Fulop no ofrece resultados fuera de unos límites dependientes de la señal, mientras que el algoritmo CWAS presenta datos para cada instante de tiempo.

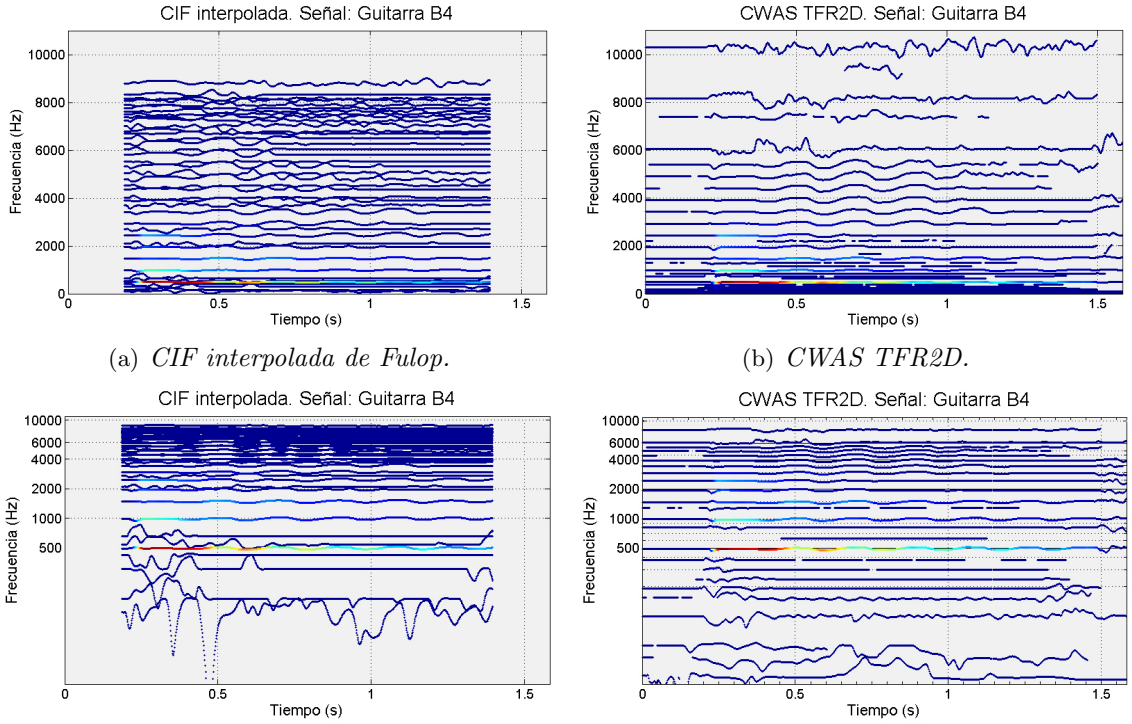
Ante representaciones tan similares, cabría preguntarse si la reconstrucción de la señal arroja asimismo resultados parecidos. A éste respecto, se ha realizado un pequeño experimento. Siguiendo de la forma más estricta posible el procedimiento para la resíntesis detallado en [58], se han sintetizado las tres señales mostradas en esta Sección. Los resultados obtenidos son, tanto numérica como sonoramente, de calidad cuestionable. Acústicamente, muestran una marcada tendencia a presentar los típicos artefactos (grillos) de interpolación en la fase. Por otro lado, aunque las envolventes de las formas de onda (sobre todo de los sonidos reales) son bastante parecidas a los originales, el valor RMS de las señales de error es del mismo orden de magnitud de la propia señal analizada (como sucede con SMS, véase Sección 5.5).

5.4.2.2. Representación 3D

Otra forma de mostrar la información disponible sería levantando en el eje Z los datos correspondientes a $A_n(t_i)$. En los trazados de tres dimensiones se pueden comparar los resultados ofrecidos por las HRSR de Fulop y el algoritmo CWAS tanto en frecuencia como en amplitud instantáneas. El resultado de esta representación aparece en las Figuras 5.22 y 5.23 para las señales del saxo y la guitarra.

Los datos ofrecidos por el algoritmo CWAS y la frecuencia instantánea canalizada interpolada de Fulop son muy similares en cuanto a forma. Sin embargo, el algoritmo CWAS ofrece los datos *reales* y puntuales de amplitud para cada parcial, mientras que la información correspondiente de Fulop debe ser tratada adecuadamente en un proceso previo a una eventual resíntesis.

Una de las ventajas más evidentes de este tipo de representación es que permite evaluar visualmente la importancia energética de los parciales. Por ejemplo, en la Figura 5.22 se



(a) *CIF interpolada de Fulop.*

(b) *CWAS TFR2D.*

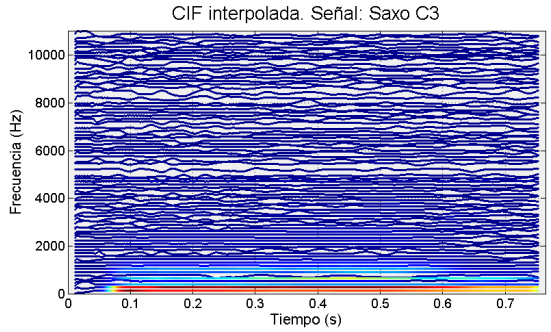
(c) *CIF interpolada de Fulop, eje frecuencial logarítmico.*

(d) *CWAS TFR2D, eje frecuencial logarítmico.*

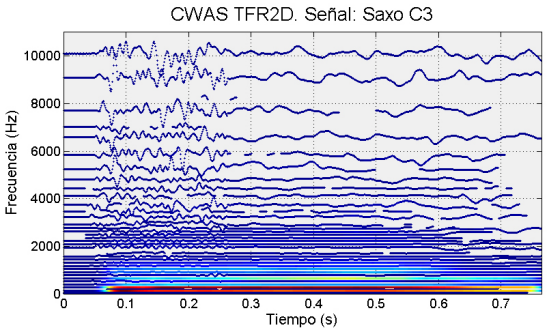
Figura 5.20: *Representaciones tiempo–frecuencia bidimensionales. Señal: Guitarra B4. (a) Frecuencia instantánea canalizada (interpolada) de Fulop. (b) Representación tiempo–frecuencia CWAS (2D). (c) y (d) Mismos datos representados en papel semilogarítmico para poder distinguir los detalles de baja frecuencia.*

puede apreciar que la mayoría de la información energética se concentra en la fundamental del saxo y sus primeros 7 armónicos (más concretamente, en los parciales 1°, 2°, 3°, 5°, 7° y 8°). Como se vio en el Capítulo 4, esto no significa que el resto de parciales no sea importante en la resíntesis. Son de hecho los responsables del color característico de la señal. Con respecto a la señal de guitarra, Figura 5.23, la concentración energética más importante se centra en la fundamental y los primeros 5 armónicos.

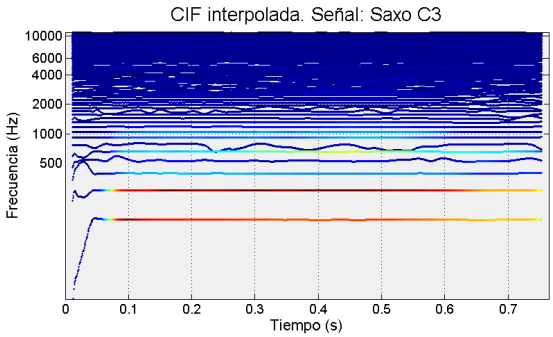
La representación tridimensional del algoritmo CWAS resulta especialmente flexible, pues permite una gran variedad en la representación gráfica. Así, se hace posible visualizar sólo los parciales energéticamente más importantes, o la parte armónica de la señal (lo cual viene a ser muchas veces lo mismo), o simplemente escoger el o los parciales que se desee sean presentados. Esta flexibilidad queda patente en la Figura 5.24, donde se presentan sucesivas representaciones de los primeros armónicos de la guitarra, Figuras 5.24(a) y 5.24(b), y



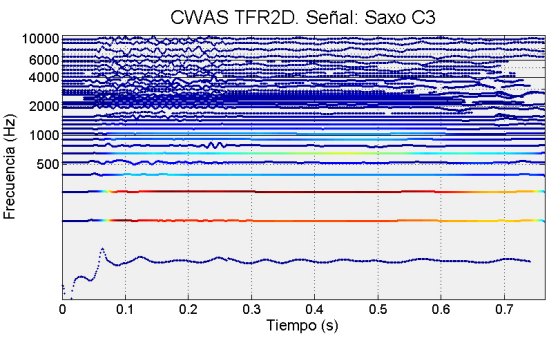
(a) *CIF interpolada de Fulop.*



(b) *CWAS TFR2D.*



(c) *CIF interpolada de Fulop, eje frecuencial logarítmico.*



(d) *CWAS TFR2D, eje frecuencial logarítmico.*

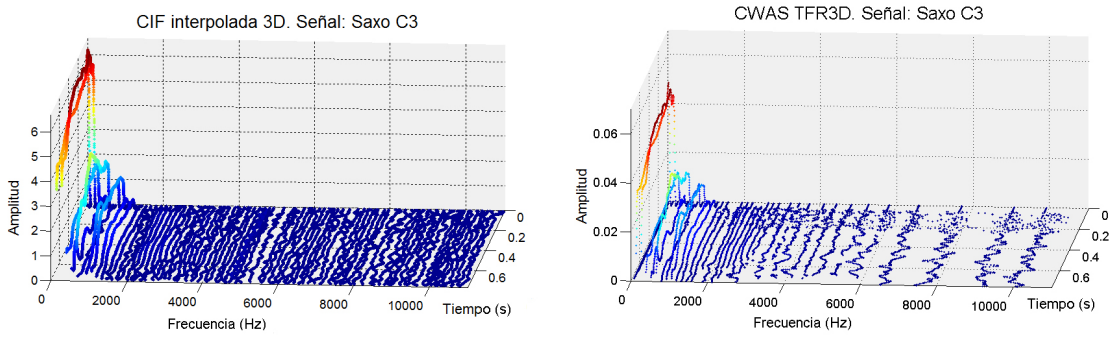
Figura 5.21: *Representaciones tiempo–frecuencia bidimensionales. Señal: Saxo C3. (a) Frecuencia instantánea canalizada (interpolada) de Fulop. (b) Representación tiempo–frecuencia CWAS (2D). (c) y (d) Mismos datos representados en papel semilogarítmico para poder distinguir los detalles de baja frecuencia.*

finalmente el parcial fundamental en solitario, Figura 5.24(c). El bending puede revelarse con gran claridad en esta representación 3D.

5.5. Síntesis de señales de audio

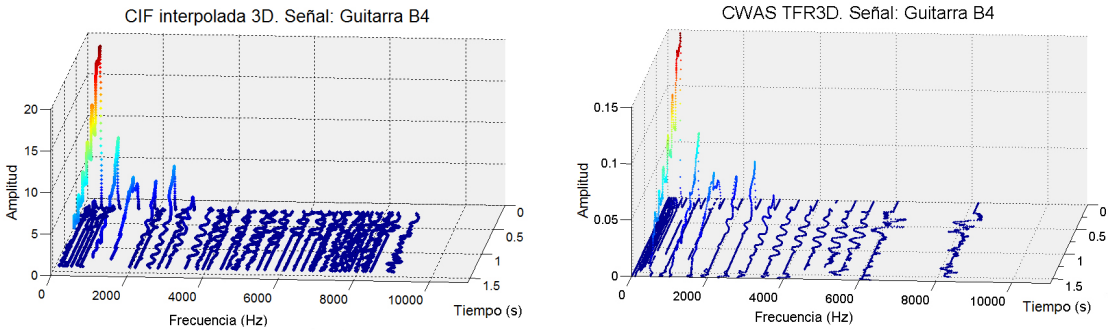
Para terminar, se va a establecer una comparación numérica y gráfica entre la calidad final de la resíntesis de señales de audio reales a través del algoritmo CWAS y reconocida técnica de la Síntesis por Modelado eSpectral (SMS).

El modelo básico y la aplicación de SMS, se ha desarrollado a partir de la tesis doctoral de Xavier Serra [148], si bien desde entonces muchas extensiones se han propuesto en el Grupo de Tecnología Musical de la Universidad Pompeu-Fabra (MTG-UPF) y por otros investigadores. SMS es un conjunto de técnicas e implementaciones de software para el



(a) *CIF interpolada de Fulop en 3D.* (b) *CWAS TFR3D.*

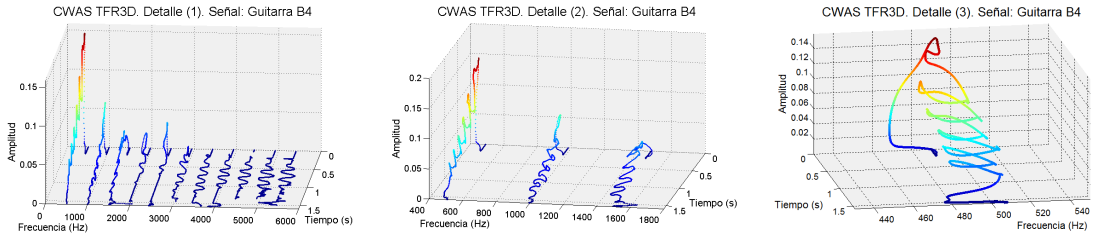
Figura 5.22: *Representación tiempo-frecuencia en 3 dimensiones de las resultados de los algoritmos HRSR y CWAS. Señal: saxo C3. (a) Representación 3D de la CIF de Fulop. (b) Representación 3D ofrecida por el algoritmo CWAS.*



(a) *CIF interpolada de Fulop en 3D.* (b) *CWAS TFR3D.*

Figura 5.23: *Representación tiempo-frecuencia en 3 dimensiones de las resultados de los algoritmos HRSR y CWAS. Señal: guitarra B4. (a) Representación 3D de la CIF de Fulop. (b) Representación 3D ofrecida por el algoritmo CWAS.*

análisis, transformación y síntesis de los sonidos musicales basado en un modelo de la señal de audio parte determinístico, parte aleatorio. Los sonidos, ya sean producidos por instrumentos musicales o por otros medios, están compuestos por un conjunto de sinusoides (parciales) y un elemento residual (ruido). Estas técnicas pueden ser utilizadas en aplicaciones de síntesis, procesamiento y codificación de señales, mientras que algunos de los resultados intermedios también podrían aplicarse a otros problemas relacionados con la música, tales como separación de fuentes, acústica y percepción musical, etc.



(a) *Fundamental y primeros 10 armónicos.* (b) *Fundamental y primeros 2 armónicos.* (c) *Parcial fundamental de la guitarra.*

Figura 5.24: *Detalles de la representación 3D del algoritmo CWAS. Señal: Guitarra B4: (a) Los 11 parciales más energéticos. (b) Los 3 parciales más energéticos. (c) Fundamental de la guitarra.*

5.5.1. Recuperación de amplitud y frecuencia instantáneas: resultados numéricos

Como se ha demostrado, el algoritmo CWAS es una herramienta muy precisa en la recuperación de la información frecuencial de la señal de audio. Las señales analizadas pueden ser fácilmente resintetizadas, lo cual permite una evaluación numérica de la precisión en la obtención de características de alto nivel. A continuación se detallan los resultados de síntesis obtenidos para cuatro de las señales anteriormente empleadas, en concreto el chirp lineal, el chirp cuadrático, el chirp exponencial y la señal de FM con excursión sinusoidal.

Los resultados numéricos correspondientes aparecen en la Tabla 5.4. Para cada caso, en la tabla se muestran los valores RMS de la señal analizada y de su respectiva señal de error, así como el error relativo, RMS_{error}/RMS_{in}, en decibelios.

Señal analizada	RMS original	RMS error	Error relativo (dB)
LC	0.6778	8.3782E-5	-77.38
QC	0.6778	4.3352E-4	-74.97
HC	0.6778	5.2107E-4	-60.93
FM	0.6778	1.0514E-5	-94.16

Tabla 5.4: *Resultados de calidad en la recuperación de la información de $A(t)$ y $f_{ins}(t)$ para las 4 señales sintéticas analizadas.*

Como se deduce del bajo valor del error, la calidad en la resíntesis es muy elevada, lo que hace que los sonidos original y sintético sean muy difíciles de distinguir. Este resultado,

como se demostrará, es extrapolable a todas las señales analizadas. Una pequeña ampliación sobre estos resultados aparece en el Anexo II.b.

5.5.2. Síntesis de sonidos reales

Como se adelantó al final del Capítulo 3, el modelo de síntesis aditiva de la señal introducido en la Sección 3.5.2, Ecuación (3.24), es el modo adoptado de resíntesis de la información.

El procedimiento por defecto ejecutado con cada una de las señales analizadas en la presente Tesis es la obtención de la señal sintética. Se han efectuado centenares de análisis bajo múltiples variantes de CWAS en cada uno de sus diferentes bloques constitutivos: utilizando diferentes bancos de filtros, técnicas de renormalización por sobrepeso o por estimación de gaussiana (ver Secciones 3.6.1 y 2.4.1.1, respectivamente), métodos de localización de picos y corte en bandas alternativos (Sección 3.8), o tracking de parciales de distintas naturalezas (Sección 3.9). A partir del momento en que se superaron los problemas iniciales del algoritmo presentado en [17], los resultados tienden a presentar un muy elevado grado de coherencia. En el presente apartado se van a resumir tres resultados seleccionados de entre los demás como representativos, mientras que otros resultados (tomados aleatoriamente de entre los numerosos análisis llevados a cabo) se expondrán en el Anexo II.c.

5.5.2.1. Resultados numéricos y figuras de mérito

De cara a reflejar la calidad final de la resíntesis, se recurre a un resultado numérico y a dos figuras de mérito por señal. En cuanto al resultado numérico, e, se trata del máximo error cometido en la resíntesis relativo a la señal original, en dB. Es decir:

$$e = 20log_{10}\left\{\frac{max\{abs[e(t)]\}}{max\{abs[x(t)]\}}\right\} \tag{5.9}$$

donde $e(t)$ es la señal de error, calculada a través de la Ecuación (3.25).

En cuanto a las figuras de mérito, se trata de las gráficas de la señal de error $e(t)$ y de su espectro. Cabe destacar que para que las señales de error resulten razonablemente visibles en una gráfica, se ha de multiplicar la señal $e(t)$ por un factor mínimo de 20.

Las señales presentadas en esta Sección son relativamente largas, pero la resíntesis es casi tan precisa como en señales más cortas y simples, como se demostrará. Se van a analizar con cierto nivel de detalle tres señales: Como ejemplo de síntesis de voz, una grabación del inconfundible Elvis Presley interpretando una estrofa del tema *Love me Tender*, concretamente la frase *"you have made my life complete"*. Como ejemplo de una señal musical armónica, un violín ejecutando una melodía, y por último, la grabación de un solo de batería

como ejemplo de una señal de transitorios muy marcados, donde el modelo de síntesis por parciales introducido parece menos intuitivo.

En la Figura 5.25, se puede apreciar en detalle las formas de onda correspondientes a este análisis. La gráfica superior se corresponde con la forma de onda de la señal original, la central con la forma de onda de la señal sintética y la inferior (línea azul) se corresponde con la señal de error, calculada a través de la Ecuación (3.25). En trazo rojo, se ha representado la señal de error multiplicada por 20. El error máximo temporal cometido para esta señal usando la Ecuación (5.9) es de -26.6761 dB.

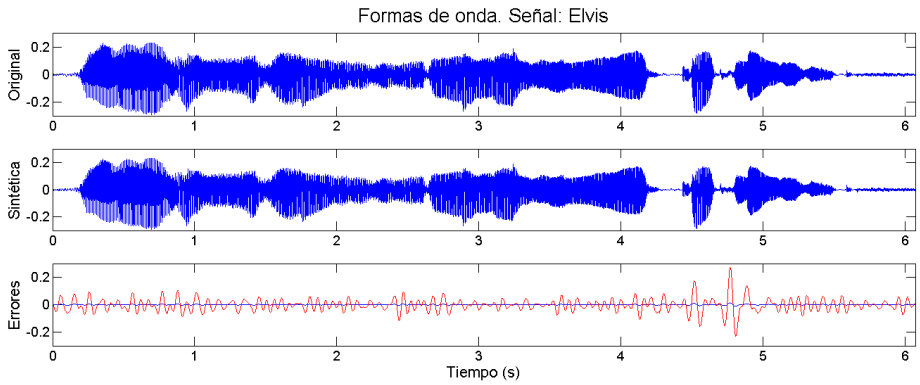


Figura 5.25: *Formas de onda. Arriba, señal original. Centro, señal sintética. Abajo, señales de error (en rojo, multiplicado por 20). Señal analizada: "Elvis".*

A continuación, en la Figura 5.26, aparecen representados los espectros de las tres señales (original, sintética y señal de error), en dB. Los espectros han sido calculados a través de la FFT. Se puede observar cómo los espectros de la señal original y sintética son prácticamente idénticos. En la gráfica inferior de esta figura, aparece el espectro de la señal de error, que tal vez sea más significativo, ya que a partir de la suavidad de la información frecuencial se deduce la falta casi absoluta de correlación de la señal de error con la señal original, de lo que se desprende el parecido entre los espectros sintético y original.

Los resultados para la señal del violín aparecen en las Figuras 5.27 y 5.28. La duración total de la señal es de 4.55 segundos, en la cual se han analizado un total de 25 frames en los que se han detectado hasta 52 parciales diferentes. Concretamente, en la Figura 5.27, han sido representadas las formas de onda original y sintética (gráficas superior y central) y en la gráfica inferior, en azul la señal de error y en trazo rojo el error aumentado (en esta ocasión multiplicado por 50 para que resulte más evidente). El máximo error temporal cometido es de -39.02 dB. En la Figura 5.28 aparecen dibujados los espectros de la señal original, sintética y de error. Se puede observar una vez más que, salvo pequeñas irregularidades

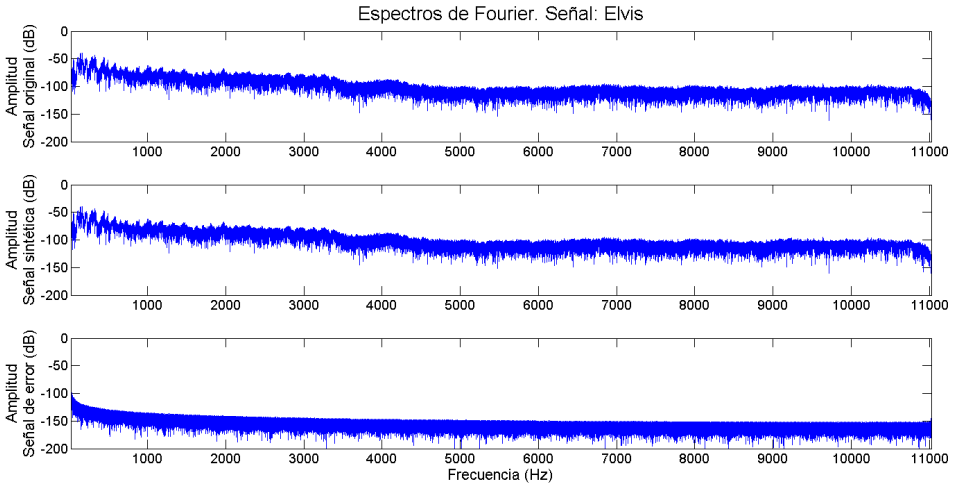


Figura 5.26: *Espectros. Arriba, señal original. Centro, señal sintética. Abajo, señal de error. Señal analizada: "Elvis".*

en la zona de muy baja y de muy alta frecuencia, los espectros original y sintético son prácticamente idénticos.

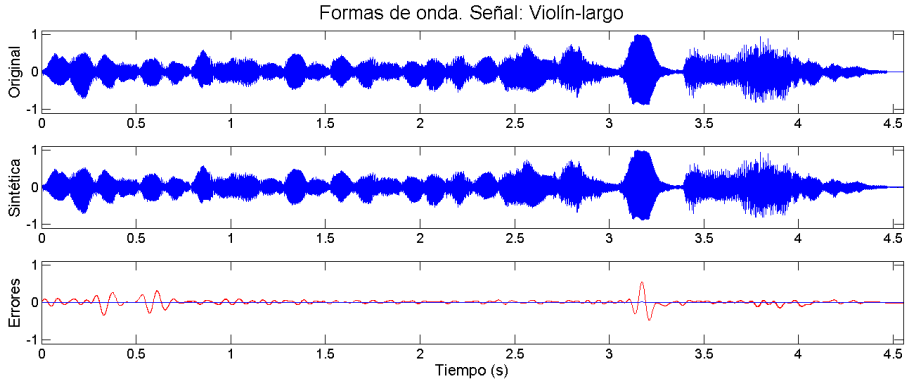


Figura 5.27: *Formas de onda. Arriba, señal original. Centro, señal sintética. Abajo, señales de error (en rojo, multiplicado por 50). Señal analizada: "Violín-largo".*

Como último ejemplo, se presentan los resultados para la señal de una batería, donde aparecen golpes rítmicos de bombo, cajas y charles. Se trata de la misma señal que será analizada en la Sección 4.5, en busca de onsets. Esta señal ha sido escogida precisamente porque el mismo tratamiento en parciales de una señal tan marcadamente transitoria como ésta

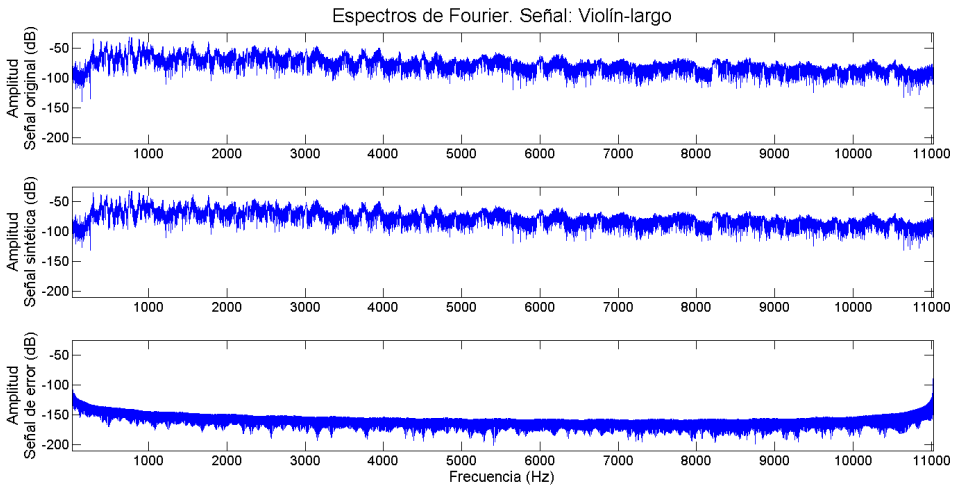


Figura 5.28: *Espectros. Arriba, señal original. Centro, señal sintética. Abajo, señal de error. Señal analizada: "Violín-largo".*

resulta poco intuitivo. Sin embargo, la calidad de los resultados obtenidos sigue siendo elevada. La duración de la señal es de 4.43 segundos, para un total de 24 frames y 54 parciales. Los resultados se muestran en las Figuras I.10 a 5.30. Pese a que el error obtenido en este caso parece más elevado que en las dos señales precedentes, el máximo error en decibelios, obtenido una vez más a través de la Ecuación (5.9) es de -27.98 dB. En la gráfica inferior de la Figura 5.29 es fácil deducir que el error (en azul, original, y en rojo, multiplicado por 20) se corresponde básicamente a un parcial de baja frecuencia, en este caso relacionado con los golpes de bombo, que no ha sido adecuadamente analizado. La misma conclusión se puede extraer de los espectros mostrados en la Figura 5.30.

Las señales sintéticas obtenidas en los tres casos, resultan difíciles de distinguir acústicamente de los originales. Salvo señales muy concretas, este resultado es generalizable a todas y cada una de las señales que han sido analizadas mediante el algoritmo propuesto. En la Sección 5.5.4 se volverá a tratar este particular desde otro punto de vista, realizando una comparativa de resultados con una herramienta clásica, concretamente el modelo SMS de Xavier Serra, [154].

5.5.2.2. Indistinguibilidad acústica

Para demostrar la calidad final de estos resultados, sería conveniente realizar pruebas de indistinguibilidad acústica con grupos aleatorios de sujetos. Esto se ha considerado innecesariamente complicado (en tiempo y recursos) dado que se puede dar una prueba indirecta

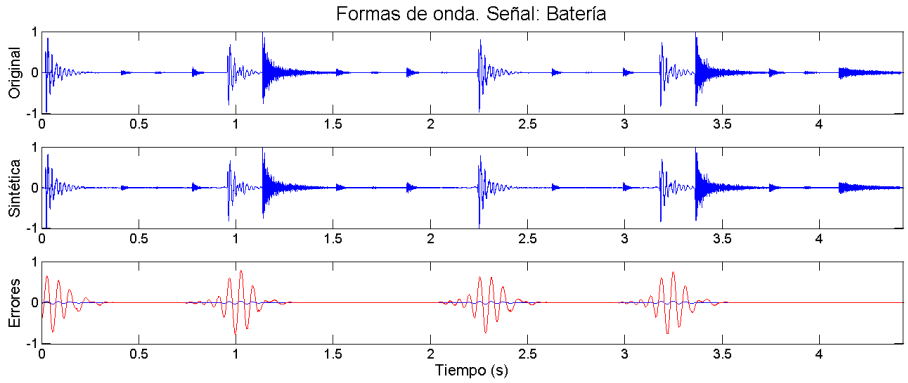


Figura 5.29: *Formas de onda. Arriba, señal original. Centro, señal sintética. Abajo, señales de error (en rojo, multiplicado por 20). Señal analizada: "Batería".*

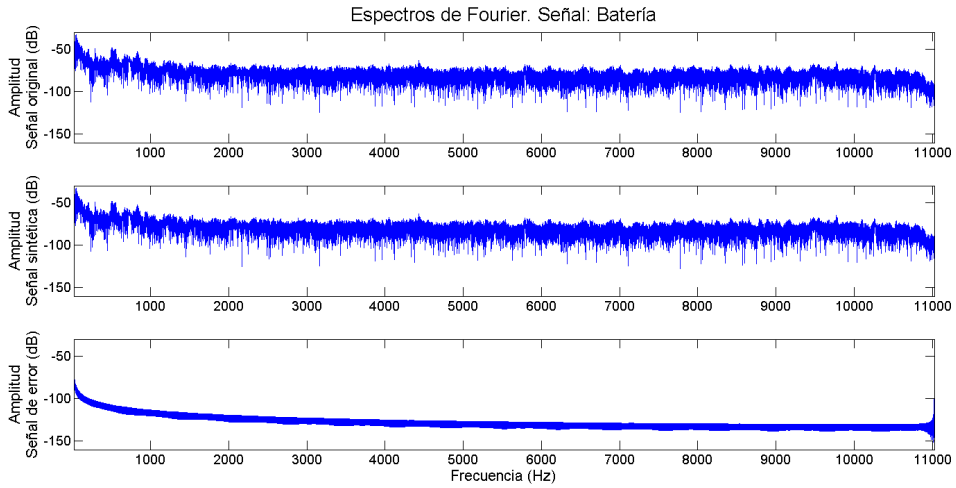
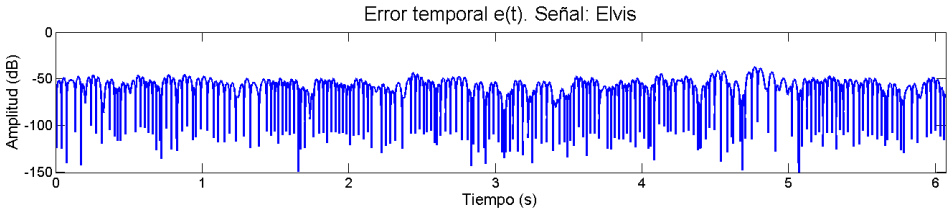


Figura 5.30: *Espectros. Arriba, señal original. Centro, señal sintética. Abajo, señal de error. Señal analizada: "Batería".*

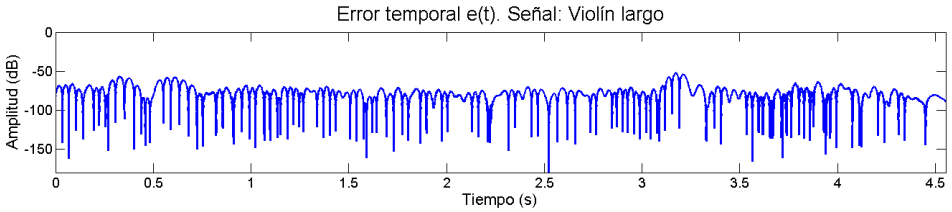
de esta cuasi–indistinguibilidad.

La prueba consiste en la presentación de los resultados de la señal de error en decibelios. En la Figura 5.31 se presentan los resultados de las tres señales analizadas, mientras que otros resultados equivalentes pueden verse en el Anexo II.c.

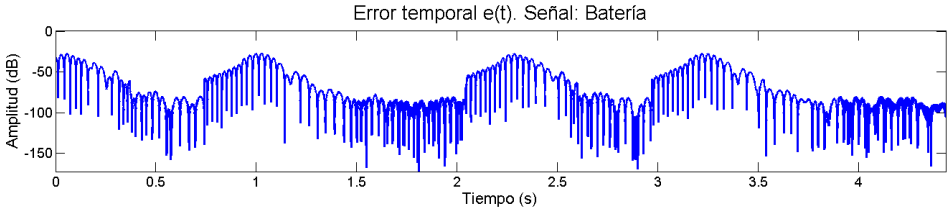
Teniendo en cuenta que el enmascaramiento del error en el proceso de digitalización (cuantificación) de una señal analógica está en torno a los -74dB, es evidente que buena parte de las señales representadas en las Figuras 5.31(b) y 5.31(c) quedan por debajo de



(a) *Error temporal $e(t)$ (dB). Señal analizada: "Elvis".*



(b) *Error temporal $e(t)$ (dB). Señal analizada: "Violín largo".*



(c) *Error temporal $e(t)$ (dB). Señal analizada: "Batería".*

Figura 5.31: *Resultados experimentales del error temporal, medido en dB. (a) Señal: "Elvis". (b) Señal: "Violín largo". (c) Señal: "Batería".*

este nivel (y por lo tanto no resultan más audibles que la diferencia entre una señal analógica y su versión grabada en, por ejemplo, un soporte CD-Audio). En cuanto a la señal de la Figura 5.31(a), si bien esta conclusión no se alcanza plenamente, está suficientemente cerca como para considerar altamente probable que una mayoría de potenciales oyentes no sean capaces de notar las diferencias. Hay que admitir que esto es una simple argumentación sin una base científica realmente sólida. Para una demostración irrefutable, no cabe otra opción que llevar a cabo una verdadera batería de pruebas de indistinguibilidad acústica.

5.5.3. Sobre el modelo SMS

La técnica SMS, de reconocido prestigio, modela un espectro dinámico como un conjunto de sinusoides (la parte determinística) definidas por una amplitud y una envolvente frecuencial lineales a tramos, más una componente ruidosa filtrada, variable en el tiempo (la parte estocástica). El diagrama de bloques de alto nivel del modelo aparece en la Figura

5.32, basada en la conferencia magistral de Xavier Serra de 2003 [151].

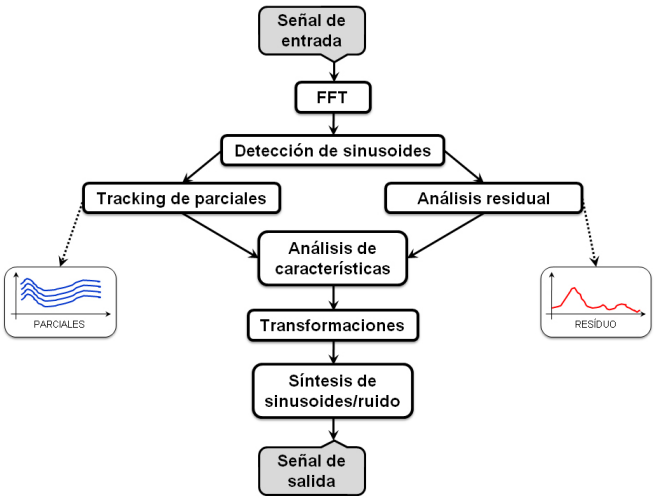
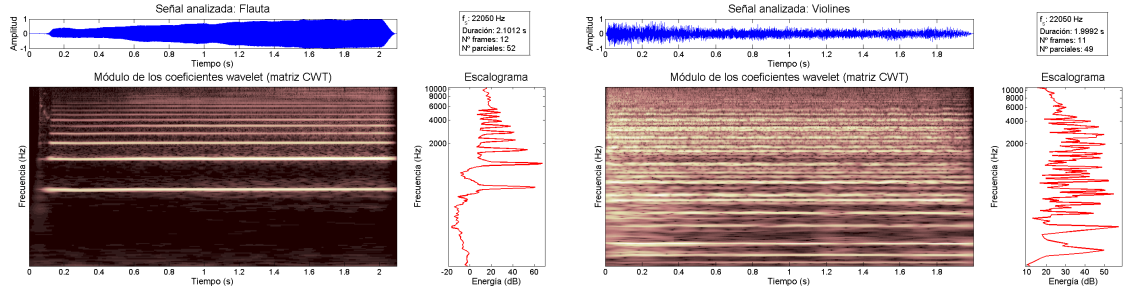


Figura 5.32: *Diagrama de bloques del algoritmo SMS.*

Como se puede apreciar en la figura, la señal de entrada es analizada mediante la FFT. La ventana del análisis se auto-optimiza (basándose en el resultado del pitch detectado). El análisis en módulo y fase de la información proporciona una serie de picos, correspondientes con los diferentes parciales detectados, que son a su vez utilizados para calcular el pitch de la señal. Cada pico espectral es localizado y seguido a medida que evoluciona, obteniéndose una amplitud y fase (de la cual se extrae la frecuencia) individuales, que caracterizan cada parcial. Tales amplitud y fase se obtienen en una muestra de cada cierto número (*hopsize*) y el resto de la información se genera interpolando los datos experimentales, pese a lo cual se pueden obtener resultados muy precisos en la amplitud y frecuencia instantáneas, aunque no así en la fase. Posteriormente, un modelo de síntesis aditiva genera la parte determinística de la señal, que es substraída a su vez de la señal original. El residuo restante es de nuevo analizado mediante la FFT, obteniéndose una aproximación espectral del mismo [149]. Sumando las partes determinística y estocástica de la señal, se sintetiza una señal de salida que conserva las características de tono y timbre de la señal original. La técnica, que como se ha dicho está en constante evolución y desarrollo por parte del MTG-UPF y otros grupos [69, 149], ha derivado eventualmente [152] en una herramienta software de análisis/síntesis de señales de audio, llamada *SMS Tools* [153] distribuida bajo Licencia Libre de Documentación GNU [153], disponible en su versión Mac OSX en la dirección `http://clam-project.org/download/mac/`.

5.5.4. Resultados numéricos

Se han analizado dos señales reales (la primera, el tono $F\#5$ de una flauta; la segunda, una pequeña orquesta de cuerdas, violines y violas principalmente, tocando una nota) a través del algoritmo CWAS (espectrogramas wavelet e información adicional en la Figura 5.33), y comparado los resultados con los obtenidos mediante el empleo de *SMS Tools*. En el análisis a través de este último se han utilizado diferentes parámetros de control, si bien los resultados finales resultan en todos los casos muy similares. Los datos aquí presentados han sido obtenidos bajo un tamaño de ventana de 2049 muestras, un tamaño de superposición (overlap) de 256, y utilizando una ventana de análisis/síntesis Blackman-Harris de 92dB (los autores del programa explican que es necesario el uso de esta ventana en el análisis de la componente residual con el fin de obtener buenos resultados).



(a) *Forma de onda normalizada, módulo de los coeficientes wavelet, escalograma total y datos relevantes de la señal "Flauta".*

(b) *Forma de onda normalizada, módulo de los coeficientes wavelet, escalograma total y datos relevantes de la señal "Violines".*

Figura 5.33: *Formas de onda, espectrogramas y escalogramas wavelet de las dos señales sintetizadas por CWAS y SMS. (a) Señal "Flauta". (b) Señal "Violines".*

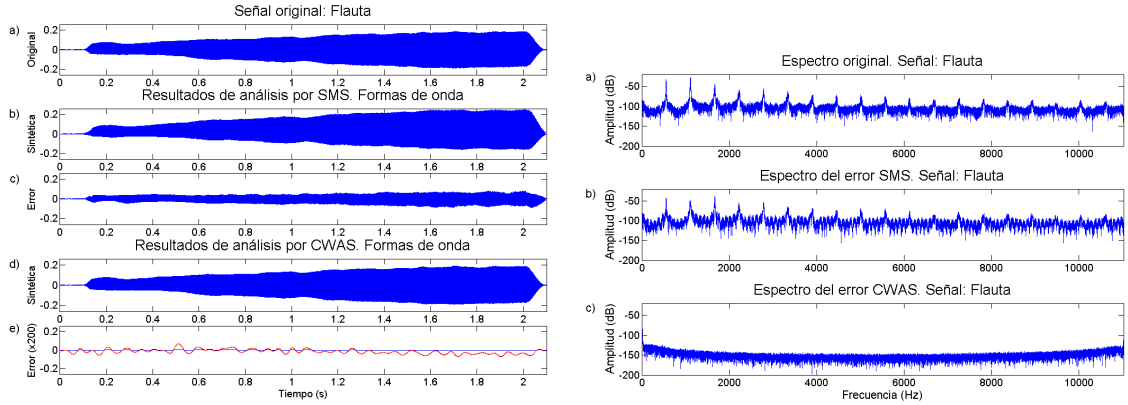
Como se viene haciendo a lo largo de éste trabajo, para demostrar la alta calidad de los resultados obtenidos, se recurrirá a mostrar tanto resultados gráficos como numéricos. Respecto a los gráficos, en el dominio temporal se van a mostrar las señales original y sintética obtenidas por SMS y CWAS y las señales de error calculadas mediante la Ecuación (3.25) para ambos métodos. En cuanto al dominio frecuencial, se presentarán para cada señal, el espectro original y los espectros de las señales de error (SMS y CWAS).

Respecto a la primera de las señales sintetizada, correspondiente a la flauta, tiene una frecuencia de muestreo $fs = 22050$ Hz y una duración aproximada de 2.1 segundos. La pureza espectral de la señal la hace especialmente útil para comparar ambas técnicas ya que, *a priori*, no deberían haber errores importantes en la interpolación de la fase del modelo de SMS. Por otro lado, la separación de las diferentes componentes espectrales proporciona una clara relación biyectiva entre los parciales detectados y los realmente existentes (en otras

palabras, las componentes que baten son prácticamente inexistentes).

En la Figura 5.33(a) aparecen representadas la forma de onda (normalizada), el espectrograma (módulo cuadrático de la matriz de coeficientes wavelet) y el escalograma total de esta señal. En el recuadro adjunto correspondiente a esta subfigura, aparecen resaltadas otras características importantes de la señal como su duración exacta y el número de parciales detectados y seguidos por el algoritmo.

Como se ha adelantado, esta señal ha sido analizada tanto por medio del algoritmo CWAS como a través de *SMS Tools*, generando sendas señales sintéticas de salida, que acústicamente resultan muy similares a la señal original. Sin embargo, los resultados numéricos y gráficos son concluyentes. En la Figura 5.34(a) se representan gráficamente, por orden, la señal original, la señal sintética y la señal de error instantáneo obtenidas empleando SMS, y la señal sintética y señales de error (en rojo, zoom 200×) obtenidas empleando CWAS.



(a) *Resultados temporales. Señal: "Flauta".* (b) *Resultados espectrales. Señal: "Flauta".*

Figura 5.34: *Resultados de síntesis de la señal "Flauta". (a) Dominio temporal: Formas de onda original, sintéticas (por SMS y CWAS) y errores (idem). (b) Dominio frecuencial: Espectro original, error por SMS y error por CWAS.*

Como se deduce de la información presentada en la Figura 5.34, el algoritmo CWAS consigue una forma de onda temporal extremadamente fiel (punto por punto) a la señal original. Tanto es así, que la señal de error debe ampliarse sensiblemente para resultar discernible de una línea plana y nula. Este hecho resalta de forma evidente la coherencia en fase de la técnica propuesta. Los resultados obtenidos mediante el uso de SMS son sensiblemente peores. Obsérvese que la señal de error tiene la misma magnitud que la señal sintética (lo que combinado con su espectro resulta en una señal de error que *suena* similar a la señal original). Respecto a la información espectral, se puede ver cómo la señal de error instantáneo SMS está claramente correlada con la señal original, mientras que el espectro

del error CWAS se limita, básicamente, a ruido de fondo.

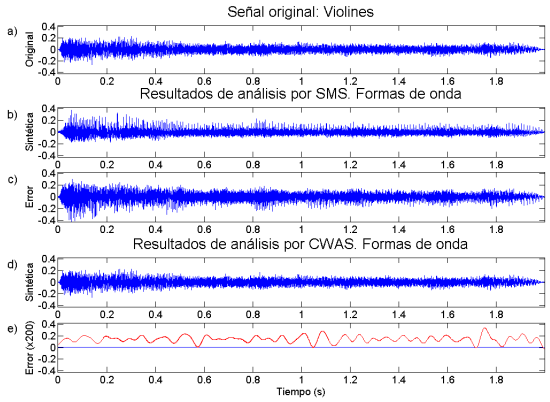
En la segunda señal, fs =22050 Hz y la duración aproximada es de 2 segundos. Los diferentes instrumentos ejecutantes (violines y violas, como se ha adelantado) presentan una marcada tendencia a batir unos con otros (efecto de coro), lo que dificulta en gran medida el proceso de seguimiento de parciales en el modelo de SMS (y por tanto la calidad final de la resíntesis). El algoritmo CWAS, por el contrario, tiende a considerar todos estos componentes batiendo como un solo parcial. Por lo tanto, se obtienen su amplitud y fase instantáneas con notable precisión, de modo que aunque el modelo subyacente no parece ser el más intuitivo (sobre todo de cara a ciertas aplicaciones), la precisión del seguimiento, y por lo tanto la calidad de la señal sintética, mejora significativamente con respecto a los resultados del modelo SMS.

En la Figura 5.33(b) aparecen representadas la forma de onda normalizada, espectrograma y escalograma total de la señal, así como sus demás características destacables. Obsérvese que se han detectado 49 parciales, menos incluso que en el caso anterior. Esto es debido, como se ha dicho, a que muchos de los parciales son el producto del coro de los diferentes instrumentos presentes, resultando fusionados en el proceso de seguimiento por parte de CWAS. Mientras tanto, SMS trata de seguir cada componente individual, perdiendo el rastro del parcial en algunas ocasiones y equivocando la trayectoria frecuencial en otras, con lo cual el resultado final presenta una variabilidad elevada, calificable casi como errática, con respecto a la señal original. Este resultado poco deseable resulta claramente audible en la señal sintética obtenida utilizando esta herramienta.

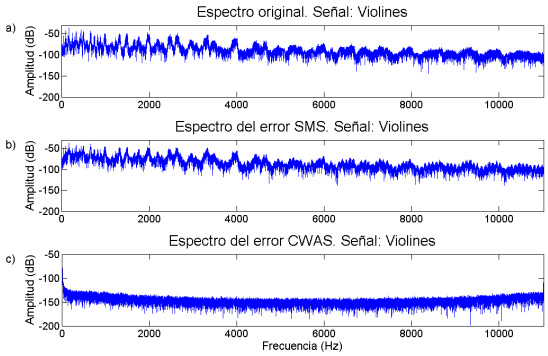
En cuanto a los resultados de la resíntesis, aparecen resumidos en la Figura 5.35. En concreto, en la Subfigura 5.35(a) se han representado, una vez más por orden, la señal original, la señal sintética y la señal de error obtenidas empleando SMS, y la señal sintética y señales de error (en rojo, zoom 200×) obtenidas empleando CWAS. En la Subfigura 5.35(b) aparecen el espectro de Fourier de la señal original, el espectro de la señal de error obtenida mediante SMS y el espectro de la señal de error sintetizada por CWAS. Las conclusiones son las mismas que en el caso anterior.

Para finalizar, en la Tabla 5.5 aparecen reflejados los valores RMS de ambas señales reales, junto con los de las respectivas señales de error obtenidas mediante las técnicas SMS y CWAS, respectivamente. La precisión en los resultados obtenidos resulta evidente a tenor de estos resultados. El algoritmo CWAS, en el peor caso, mejora en más de 11dB el resultado obtenido utilizando SMS.

Respecto a la validez del modelo subyacente (al menos en lo que al concepto de parcial se refiere), parece que la técnica clásica de SMS, basada en la FFT, resulta más intuitiva (sobre todo con señales como la del segundo ejemplo, donde el batido del coro de cuerdas *abre* en frecuencia el parcial subyacente a CWAS, ampliando el concepto a su nivel más abstracto).



(a) *Resultados temporales. Señal: "Violines".*



(b) *Resultados espectrales. Señal: "Violines".*

Figura 5.35: *Resultados de síntesis de la señal "Violines". (a) Dominio temporal: Formas de onda original, sintéticas (por SMS y CWAS) y errores (idem). (b) Dominio frecuencial: Espectro original, error por SMS y error por CWAS.*

Señal analizada	RMS original	SMS error (dB)	CWAS error (dB)
Flauta	0.0805	-8.92	-49.73
Violines	0.0442	2.38	-34.11

Tabla 5.5: *Resultados de calidad en la recuperación de la información para las dos señales reales analizadas.*

De cara a que los resultados de SMS fuesen sonora y numéricamente comparables a los del algoritmo CWAS, se haría necesario un análisis de la información punto por punto, similar al presentado en la Sección 5.2.1. En tal caso, independientemente de las comparaciones en la calidad de la resíntesis que pudiesen obtenerse, el algoritmo CWAS resultaría, como se ha demostrado, significativamente más rápido que SMS [120].

5.6. Conclusiones y contribuciones

Como se ha demostrado a lo largo de este Capítulo, el algoritmo CWAS, en comparación con otras herramientas de análisis y síntesis de señales de audio, es una herramienta rápida, precisa y muy manejable.

Como se ha visto, los tiempos de computación del algoritmo desarrollado en comparación

con técnicas basadas en la STFT son sensiblemente menores, bajo condiciones de trabajo equivalentes. Concretando, el algoritmo CWAS viene a ser al menos entre 6 y 9 veces más rápido, dependiendo del equipamiento informático utilizado. La ventaja tiende a crecer a medida que las señales analizadas son más largas (ratios entre 14 y 6.5).

En cuanto a la precisión en la recuperación de la frecuencia instantánea, el algoritmo CWAS se ha mostrado, asimismo, una herramienta más eficiente. En el conjunto de señales sintéticas analizado, se puede ver que los resultados son siempre mejores empleando el algoritmo presentado en esta disertación, si bien la precisión obtenida por otros métodos (relacionados con la STFT y la WVD) es, en muchos casos, suficientemente elevada.

Por lo que se refiere a la precisión en la resíntesis de la señal, el algoritmo CWAS presenta unos resultados sensiblemente mejores que, en este caso, la técnica de SMS. Es ésta precisión en la representación temporal de la señal (y por ende de cada uno de sus parciales constitutivos detectados) la que lleva asociado el uso potencial de la información obtenida para múltiples aplicaciones (véanse el Capítulo 4 y el Anexo).

Por último, se ha comparado en diferentes formas la representación de la información arrojada por el algoritmo desarrollado en comparación con representaciones extraídas de la Time-Frequency Toolbox de Auger, y con el algoritmo de Reasignación de Fulop [64] y sus espectrogramas asociados. Como se ha visto, la calidad visual de la información es sensiblemente mejor en el algoritmo CWAS, que además permite una visualización bi y tridimensional de las componentes de la señal altamente flexible.

Aparte de los análisis comparativos obtenidos, las contribuciones presentadas en este Capítulo, han sido:

1. Algoritmo de síntesis de sonidos.
2. Representación Tiempo–Frecuencia basada en CWAS.
   - 2–D.
   - 3–D.

Capítulo 6

Conclusiones y Líneas de Trabajo

"Mientras estés solucionando un problema,
no te preocupes. Después de que
lo hayas resuelto, no obstante, llega
el momento de preocuparse".
Richard Phillips Feynman (1918-1988).
Físico estadounidense.

En esta disertación se ha presentado un algoritmo funcional basado en la Transformada Wavelet Continua y Compleja bautizado como algoritmo de Síntesis Aditiva por Wavelet Complejas, CWAS por sus siglas en inglés. El algoritmo parte de una versión previa (presentada en 2003, [17]), cuyas limitaciones prácticas ponían en tela de juicio la aplicación práctica de la transformada.

Sin embargo, un análisis matemático riguroso de los coeficientes wavelet (obtenido para una serie de señales algebraicamente resolubles de modo exacto o con un grado de aproximación suficiente) ha puesto en evidencia cuáles eran las naturalezas de tales limitaciones (Capítulo 2). Tras encontrar la solución para estos problemas, se ha llegado finalmente a proponer un nuevo modelo de la señal de audio (Capítulo 3) situado en la mejor de las posiciones de cara a posibles aplicaciones (Capítulos 4 y 5). El algoritmo CWAS resulta ser una rápida y eficaz herramienta de extracción de características de alto nivel de la señal de audio (Capítulo 5).

A través de un filtrado pasobanda complejo unitario se obtienen los coeficientes wavelet, cuyo módulo proporciona de forma indirecta las bandas del espectro de frecuencia influenciadas por la misma componente. La suma de los coeficientes wavelet en estas bandas proporciona una función compleja para cada parcial detectado cuyas amplitud y fase instantáneas son altamente coherentes, en el sentido de que quedan muy cerca del par canónico

teórico de cada parcial de la señal. Un seguimiento de tales parciales a lo largo del tiempo, en una estructura frame-to-frame anidada (Capítulo 3) permite el análisis de señales de duración arbitraria, así como el acceso a la información de cada parcial detectado por parte del usuario.

La gran ventaja del modelo propuesto consiste en esta coherencia de fase. Al conocerse de forma muy precisa amplitud y fase instantáneas (y por lo tanto f_{ins}) de cada parcial de la señal en cada instante de tiempo, un simple modelo de síntesis aditiva permite la generación de una señal sintética que no sólo presenta las mismas características tímbricas y de pitch de la señal original, sino que, punto por punto, la diferencia entre las formas de onda de las señales analizada y sintética resulta despreciable para la mayoría de las aplicaciones (Capítulo 5).

El tiempo de procesado requerido en entorno Matlab® (actualmente situado entre 10× y 15× respecto a la duración de la señal, analizada en un equipo de altas prestaciones), por encima del requerido por la STFT en sus condiciones normales de trabajo, es la limitación principal de la técnica propuesta. Esto, que parece una limitación importante, está sin embargo muy lejos de serlo. De hecho, la coherencia de fase no resultaría una ventaja tan evidente sobre otros modelos de Distribuciones Tiempo–Frecuencia, como las basadas en la STFT, si no se la añadiese el importante dato de que el algoritmo CWAS es sensiblemente más rápido que la STFT *en condiciones de trabajo equivalentes* (concretamente, como se ha demostrado experimentalmente, entre 9 y 14 veces más rápido en equipos de sobremesa de prestaciones elevadas, Capítulo 5). Del mismo modo, la precisión en la recuperación de la frecuencia instantánea de la señal es mejor en el algoritmo CWAS que en herramientas basadas tanto en la STFT (HRSR) como en la PWVD, incluyendo sus versiones reasignadas.

La versatilidad del modelo subyacente permite colocar al algoritmo CWAS en buena posición de cara a múltiples aplicaciones. De hecho, se ha utilizado en síntesis de sonidos, localización de onsets y detección de fundamentales, siempre con resultados muy prometedores (Capítulos 4 y 5). Del mismo modo, se han hecho grandes progresos de cara a aplicaciones más completas y complejas, como la separación ciega de fuentes monaurales de sonido (Capítulo 4).

Por lo tanto, se ha demostrado que la Transformada Wavelet Continua y Compleja puede ser una herramienta precisa para la obtención de características de alto nivel de la señal de audio, siendo el algoritmo CWAS una versión generalista del modelo, que habrá que mejorar tanto de cara a sus aplicaciones concretas como para su implementación hardware.

En este sentido, se abren cinco posibles líneas de investigación:

1. La primera de ellas es continuar con lo expuesto en el Capítulo 4 de esta Tesis, mejorando el algoritmo de separación de parciales superpuestos, adecuando el uso de los tiempos de onset y offset de las notas musicales y creando un algoritmo de detección

de timbre. Técnicas más recientes [79], proponen una evolución del principio de CAM para la estimación de envolventes, la Similitud de Envolvente Temporal Armónica (Harmonic Temporal Envelope Similarity, HTES), la cual, empleando la información de parciales no superpuestos de las notas de un instrumento determinado (ocurran cuando ocurran en la grabación), construye un modelo del instrumento que puede ser utilizado para reconstruir estimaciones de otros parciales armónicos compartidos, permitiendo incluso la separación de notas completamente superpuestas. La mezcla de todas estas técnicas podría dar lugar a la creación de un algoritmo de separación monaural muy robusto y general, sin por ello ceder demasiado en su concepción de ciego. Una de sus posibles aplicaciones sería la obtención automática de partitura para piezas polifónicas grabadas en un solo canal, además de amplificación inteligente de señales (para su eventual uso, por ejemplo, en audífonos). Las complicaciones de esta futura técnica resultan más que suficientes para la elaboración de otra Tesis doctoral. En cuanto a la separación estereofónica, se han hecho tímidas incursiones en el tema que posibilitan el lanzamiento de una nueva linea de estudio en esta dirección, con bastantes aplicaciones futuras.

2. La segunda línea principal de investigación, ya activa, consiste en la adecuación del método propuesto para otras ramas del procesado de señal diferentes del audio. En estos momentos hay dos proyectos activos: la aplicación del algoritmo CWAS para la extracción de características de señales biomédicas (ECG), y geofísicas (detección precisa de los tiempos de llegada de las ondas S, P, Love y Rayleigh).

3. La tercera línea propuesta consiste en la implementación de una versión optimizada del algoritmo CWAS para Matlab® en una máquina de procesamiento en paralelo, de cara a posibles utilizaciones de la técnica en tiempo real o con retardos suficientemente bajos como para que el abanico de posibles usos se amplíe. En cualquier caso, dada la evolución en la potencia de computación, el algoritmo propuesto podría resultar suficientemente rápido en un lapso de tiempo relativamente breve.

4. Algunas de las limitaciones en la velocidad de proceso son debidas al entorno de trabajo propio de Matlab®. En las últimas fechas se han llevado a cabo los primeros ensayos de implementación del modelo de la señal propuesto en esta Tesis en entornos de programación diferentes (concretamente mediante la programación de algoritmos en la GPU). Los resultados iniciales indican que el tiempo de computación necesario se reduce de forma considerable, por lo que cabe la posibilidad de desarrollar aplicaciones más completas que puedan trabajar prácticamente en tiempo real.

5. Por último, destaca la necesidad de estudiar la implementación electrónica de la técnica presentada, de cara a potenciales empleos del hardware asociado (como los citados

audífonos, o simplemente un nuevo sintetizador de instrumentos musicales). En la versión expuesta en esta disertación, no existe ninguna familia de sistemas hardware en el mercado que sean capaces de procesar semejante volumen de información de forma eficiente. Sin embargo, es evidente que el algoritmo no está optimizado, y que se podrían recortar algunas de sus actuales prestaciones de cara a conseguir que el sistema pueda ser programado dentro de una FPGA o similar, resultando de este modo portable y ampliando sus potenciales usos en el tratamiento de la señal musical.

Jesús Ponce de León Vázquez
jponce@unizar.es
Departamento de Ingeniería Electrónica y Comunicaciones
Universidad de Zaragoza
Zaragoza, 7 de Mayo de 2012.

Departamento de
Ingeniería Electrónica
y Comunicaciones
Universidad Zaragoza

Universidad de Zaragoza
Departamento de Ingeniería Electrónica y Comunicaciones

Análisis y Síntesis de Señales de Audio a Través de la Transformada Wavelet Continua y Compleja:

El Algoritmo C.W.A.S.

ANEXOS

Zaragoza España

Anexo I

Generalidades

"Me sorprende con qué frecuencia
se corrigen las predicciones
en los resultados experimentales".
Murray Gell-Mann (1929).
Físico teórico y profesor
estadounidense.

En este primer Apéndice se ampliarán los resultados concernientes a los tres primeros Capítulos de esa Tesis, así como cuestiones de carácter general.

I.a. Tabla de notas musicales

Aunque no ha lugar a equivocaciones desde el momento en que en el texto se define que la frecuencia de la nota musical $A4$ es de 440 Hz, en la Tabla I.1 se presentan las frecuencias de todas las notas musicales afinadas del espectro auditivo.

Las frecuencias de la Tabla I.1 obedecen a la siguiente expresión aproximada:

$$f = f_0 \cdot e^{\ln 2\left[(o-4)+\frac{n-10}{12}\right]} \tag{I.1}$$

donde $f_0 = 440Hz$ $(A4)$, $o \in [0, 9]$ es el número de octava y $n \in [1, 12]$ el número de nota, desde $DO = C = 1$ hasta $SI = B = 12$.

La Tabla I.1 será usada como referencia a lo largo del presente trabajo. Uno de los apartados en los que más se ha utilizado la información contenida es en la Sección 4.6.3, acerca de la detección de frecuencias fundamentales.

		Número de Octava									
Notas		**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
DO	**C**	16.36	32.72	65.44	130.88	261.76	523.52	1047.04	2094.08	4188.16	8376.32
DO#	**C#**	17.33	34.66	69.32	138.64	277.28	554.56	1109.12	2218.24	4436.48	8872.96
RE	**D**	18.36	36.72	73.44	146.88	293.76	587.52	1175.04	2350.08	4700.16	9400.32
RE#	**D#**	19.45	38.9	77.8	155.6	311.2	622.4	1244.8	2489.6	4979.2	9958.4
MI	**E**	20.60	41.21	82.42	164.84	329.68	659.36	1318.72	2637.44	5274.88	10549.76
FA	**F**	21.8	43.6	87.2	174.4	348.8	697.6	1395.2	2790.4	5580.8	11161.6
FA#	**F#**	23.12	46.25	92.5	185	370	740	1480	2960	5920	11840
SOL	**G**	24.5	49	98	196	392	784	1568	3136	6272	12544
SOL#	**G#**	25.95	51.91	103.82	207.64	415.28	830.56	1661.12	3322.24	6644.48	13288.96
LA	**A**	27.5	55	110	220	**440**	880	1760	3520	7040	14080
LA#	**A#**	29.13	58.26	116.52	233.04	466.08	932.16	1864.32	3728.64	7457.28	14914.56
SI	**B**	30.86	61.72	123.44	246.88	493.76	987.52	1975.04	3950.08	7900.16	15800.32

Tabla I.1: *Frecuencias fundamentales (Hz) para la escala temperada con base $LA4 = 440Hz$.*

I.b. Términos de intermodulación: análisis en escala

En la Sección 2.3.3, después de la expresión explícita de los coeficientes wavelet para n parciales:

$$\begin{aligned}\|W_x(a,b)\|^2 \approx &\sum_{i=1}^{n} A_i^2 e^{-\sigma^2(\omega_i a-\omega_0)^2} + \\ &+ \sum_{i\neq k=1}^{n} A_i e^{-\frac{\sigma^2}{2}(\omega_i a-\omega_0)^2} cos(\omega_i b)\cdot A_k e^{-\frac{\sigma^2}{2}(\omega_k a-\omega_0)^2} cos(\omega_k b)\end{aligned} \tag{I.2}$$

se introducía el concepto de términos de intermodulación:

$$A_i e^{-\frac{\sigma^2}{2}(\omega_i a-\omega_0)^2} cos(\omega_i b)\cdot A_k e^{-\frac{\sigma^2}{2}(\omega_k a-\omega_0)^2} cos(\omega_k b) \tag{I.3}$$

términos que, como se puede ver, afectan a los parciales $i-simo$ y $k-simo$ de la señal, con $i \neq k$. Sea a_i el parámetro de escala $i-simo$:

$$a_i = \frac{\omega_0}{\omega_i} \tag{I.4}$$

Por otro lado, se supondrá que el filtro de análisis es analítico, es decir:

$$\sigma^2\omega_0^2 > 25 \tag{I.5}$$

Con estas asunciones, es evidente que, en la Ecuación (I.3), se tiene:

$$\lim_{|\omega_i-\omega|\to 0} A_i e^{-\frac{\sigma^2}{2}(\omega_i a-\omega_0)^2} = \lim_{|\omega_i-\omega|\to 0} A_i e^{-\frac{\sigma^2\omega_0^2}{2\omega^2}(\omega_i-\omega)^2} = 1 \quad \forall\omega_i \tag{I.6}$$

y:

$$\lim_{|\omega_i-\omega|\to +\infty} A_i e^{-\frac{\sigma^2}{2}(\omega_i a-\omega_0)^2} = \lim_{|\omega_i-\omega|\to +\infty} A_i e^{-\frac{\sigma^2\omega_0^2}{2\omega^2}(\omega_i-\omega)^2} = 0 \quad \forall\omega_i \tag{I.7}$$

A continuación, se va a detallar matemáticamente el proceso que desemboca en la conclusión que culmina la Sección nombrada (a saber, que los términos de intermodulación resultan despreciables cuando las frecuencias involucradas están suficientemente alejadas entre sí, pero de valor elevado en caso contrario), calculando el valor de los términos de intermodulación en los casos extremos de frecuencias muy separadas entre sí y de dos frecuencias extremadamente próximas. Es obvio que en el caso de una señal de audio real se pueden encontrar, en general, parciales que se comporten según ambas posibilidades.

I.b.1. Separación interfrecuencial elevada

En este caso, supóngase que las n frecuencias involucradas, ordenadas de menor a mayor, son:

$$\omega_1 < \omega_2 < \ldots < \omega_i < \ldots < \omega_{n-1} < \omega_n \tag{I.8}$$

y están separadas dos a dos por $\Delta \to \infty$:

$$\omega_{i+1} = \omega_i + \Delta \tag{I.9}$$

De las Ecuaciones (I.6) y (I.7) se deduce que las exponenciales de los términos de intermodulación relacionadas con los parciales $i - simo$ y $k - simo$ son importantes sólo para escalas próximas a a_i o a_k, respectivamente. Como, por la Ecuación (I.9)es evidente que ω_i y ω_k están separadas por, al menos, $\Delta \to \infty$, al menos uno de los dos términos exponenciales tiende a cero mientras el otro permanece acotado por uno, y por lo tanto:

$$A_i e^{-\frac{\sigma^2}{2}(\omega_i a-\omega_0)^2} cos(\omega_i b) \cdot A_k e^{-\frac{\sigma^2}{2}(\omega_k a-\omega_0)^2} cos(\omega_k b) \to 0 \quad \forall i,k \tag{I.10}$$

I.b.2. Separación interfrecuencial despreciable

Supóngase que existe algún k para el que se tenga:

$$\omega_k = \omega_i + \Delta \tag{I.11}$$

con $|\Delta| \to 0$. En este caso, aplicando la Ecuación (I.6) a la Ecuación (I.3), los términos de intermodulación quedan:

$$A_i e^{-\frac{\sigma^2}{2}(\omega_i a-\omega_0)^2} cos(\omega_i b) \cdot A_k e^{-\frac{\sigma^2}{2}(\omega_k a-\omega_0)^2} cos(\omega_k b) \approx A_i A_k \cos(\omega_i b)\cos(\omega_k b) \tag{I.12}$$

Por lo tanto, cuando el banco de filtros analice dos de estos parciales próximos, la Ecuación (I.2) evaluada en la escala $a_1 \approx a_2$ ofrecerá como resultado:

$$\begin{aligned} \|W_x(a_1,b)\|^2 &\approx A_1^2 + A_2^2 + 2A_1A_2\cos(\omega_1 b)\cos(\omega_2 b) \Rightarrow \\ &\Rightarrow \|W_x(a_1,b)\| \in \left[|A_1 - A_2|, A_1 + A_2\right] \end{aligned} \tag{I.13}$$

El factor 2 que aparece en la Ecuación (I.13) proviene del hecho de que por cada par de parciales i y k aparecen dos términos de intermodulación cruzados, ambos iguales (i con k y k con i). Por lo tanto, los posibles términos de intermodulación del parcial $i-simo$ son en general del orden de magnitud de la propia amplitud de la señal A_i, oscilando además con una frecuencia marcada por ω_i y Δ. Estos términos de intermodulación se hacen perfectamente aparentes en la Figura I.1, así como en las Figuras 2.5 y 2.6, Sección 2.3.3.

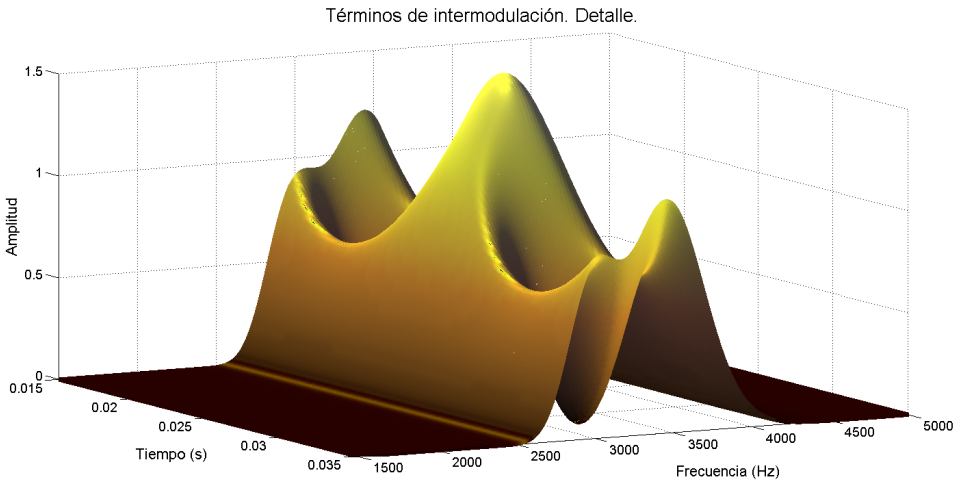


Figura I.1: *Detalle de los términos de intermodulación para una señal compuesta por dos cosenos de frecuencias muy próximas (3kHz y 3.5kHz). La proximidad entre frecuencias depende del ancho del banco de análisis. En este caso $\sigma = 0.6148$.*

I.c. Renormalización: estudio detallado

En la Sección 2.4.1, donde se trataba de los problemas relacionados con el paso de las variables continuas a variables discretas, se proporcionaron los datos experimentales de amplitud recuperada tras el análisis de señales de diferentes frecuencias bajo el algoritmo original. Como se puede ver en la Tabla I.2 (idéntica a la Tabla 2.1), las señales de entrada tienen todas amplitud 1, pero el resultado del análisis indica en ocasiones una amplitud ligeramente menor.

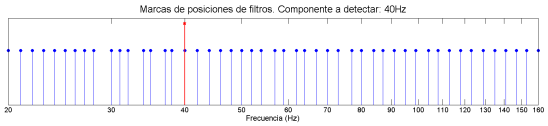
Frecuencia	**Amplitud original**	**Amplitud detectada**	**Ratio**
40	1.0000	0.9873	1.0129
80	1.0000	0.9989	1.0011
160	1.0000	1.0000	1.0000
320	1.0000	1.0000	1.0000
640	1.0000	0.9299	1.0753
1280	1.0000	0.9299	1.0753
2560	1.0000	1.0000	1.0000
5120	1.0000	0.9767	1.0238
10240	1.0000	0.9328	1.0721
20480	1.0000	0.9553	1.0468

Tabla I.2: *Resultados empíricos iniciales del análisis en amplitud de una señal de amplitud constante.*

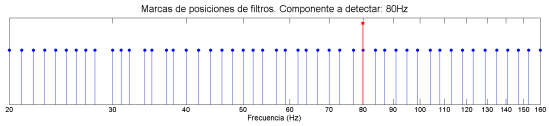
El algoritmo original [17], buscaba el mejor banco de filtros para cubrir adecuadamente el espectro de la señal de entrada. Por lo tanto, cada una de las diferentes señales analizadas tiene su propio banco de filtros de análisis asociado. En las Figuras I.2(a) a I.2(j) aparecen reflejadas la frecuencia buscada en cada caso (marcada con una estrella roja en cada figura) y la posición de las frecuencias centrales de los filtros de análisis en cada caso (tan sólo los situados en la vecindad, por propósitos de claridad).

Como se deduce de las figuras, las frecuencias de 40, 80, 160, 320, y 2560 Hz tienen un filtro centrado exactamente sobre la frecuencia buscada. De ellas, salvo de las 2 primeras, de las demás se recupera la amplitud de forma perfecta.

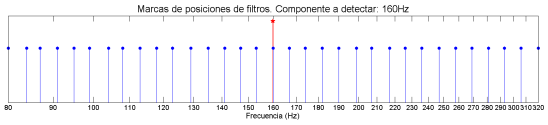
Nótese que las frecuencias de 40 y 80 Hz tienen un filtro situado exactamente encima, si bien esto no significa que puedan ser detectadas de forma ideal. En efecto, los filtros de frecuencias más bajas son calculados en un conjunto de puntos muy limitado y esto provoca la pérdida de precisión en la recuperación de la señal. Este hecho se refleja en las Figuras I.3(a) y I.3(b), cuyas representaciones gráficas son sendas ampliaciones de las estructuras de los bancos de filtros representados en el Capítulo 3, Figuras 3.2 y 3.3.



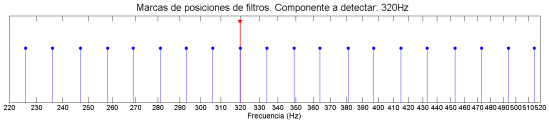
(a) *Localización de los filtros del banco. Caso 1: señal de 40 Hz.*



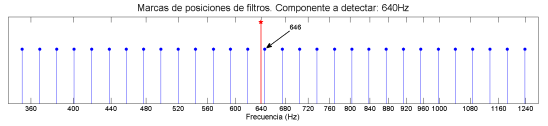
(b) *Localización de los filtros del banco. Caso 2: señal de 80 Hz.*



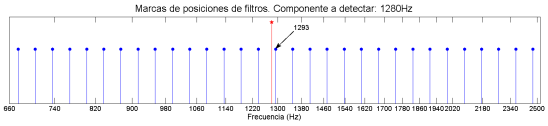
(c) *Localización de los filtros del banco. Caso 3: señal de 160 Hz.*



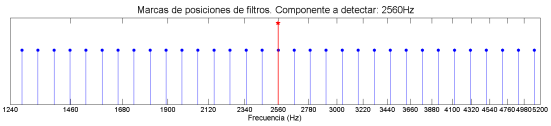
(d) *Localización de los filtros del banco. Caso 4: señal de 320 Hz.*



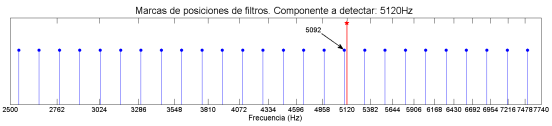
(e) *Localización de los filtros del banco. Caso 5: señal de 640 Hz.*



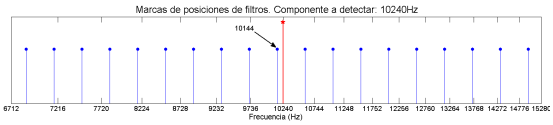
(f) *Localización de los filtros del banco. Caso 6: señal de 1280 Hz.*



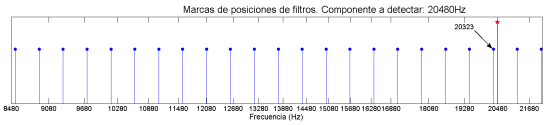
(g) *Localización de los filtros del banco. Caso 7: señal de 2560 Hz.*



(h) *Localización de los filtros del banco. Caso 8: señal de 5120 Hz.*



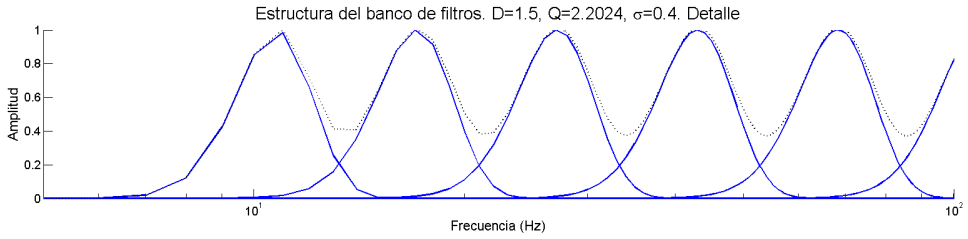
(i) *Localización de los filtros del banco. Caso 9: señal de 10240 Hz.*



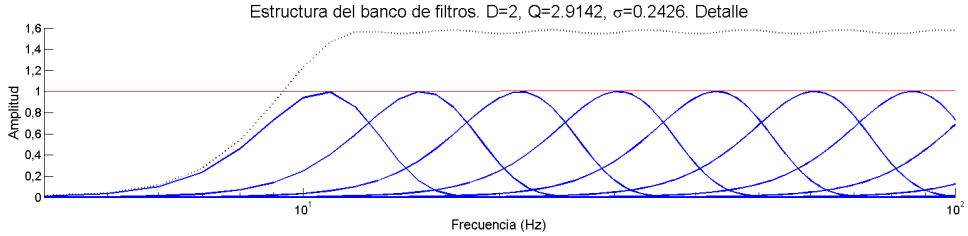
(j) *Localización de los filtros del banco. Caso 10: señal de 20480 Hz.*

Figura I.2: *Localización de los filtros del banco para diferentes tonos a localizar. La diferencia entre el tono en cuestión y el filtro más cercano es un indicador del valor máximo de la detección, tanto más bajo cuando mayor sea esta diferencia.*

En estas figuras (que no se corresponden con ninguno de los bancos de filtros de ninguno de los análisis efectuados en este punto) se puede observar como los filtros centrados en frecuencias bajas se conocen en relativamente pocos puntos, y por eso pierden su forma teóricamente gaussiana. Esto causa que algunos filtros no lleguen a alcanzar su amplitud máxima teórica (en este caso 1) y que por lo tanto algunas frecuencias no puedan ser estudiadas apropiadamente. Esta sería la causa los errores en las frecuencias de 40 y 80Hz.



(a) *Banco de filtros de cobertura incompleta. Detalle.*



(b) *Banco de filtros de cobertura completa. Detalle.*

Figura I.3: *Detalle de la estructura de bancos de filtros: (a) De cobertura incompleta. (b) De cobertura completa. En ambas figuras, la línea punteada es la contribución global.*

Los filtros de frecuencias mayores se conocen en un conjunto de puntos creciente, y por lo tanto este hecho no supone ningún problema.

I.d. Ejemplos de espectrogramas y escalogramas wavelet

El algoritmo CWAS ha sido puesto a prueba en el análisis de cientos de señales de audio diferentes, tanto sintéticas como grabaciones de instrumentos reales e incluso extractos de temas musicales. Los resultados en cuanto a la calidad de la resíntesis son excepcionales en todos los casos, con valores de error en la resíntesis por debajo de 25dB en todos los casos, y en gran parte de las señales por debajo de 50dB. Esto significa que las diferencias numéricas entre la señal real y la sintética son prácticamente despreciables, mientras que las diferencias sonoras son prácticamente inaudibles.

Como resultados de salida por defecto, el algoritmo CWAS ofrece el espectrograma y el escalograma wavelet, la forma de onda de la señal analizada (normalizada o no, por elección del usuario), y una gráfica del error instantáneo de resíntesis (en dB).

A continuación se reproducen los resultados de varias señales de audio analizadas por CWAS, en los que además se incluyen los datos técnicos más relevantes del análisis:

Es un conjunto de señales que cubre razonablemente el amplio abanico de posibilidades

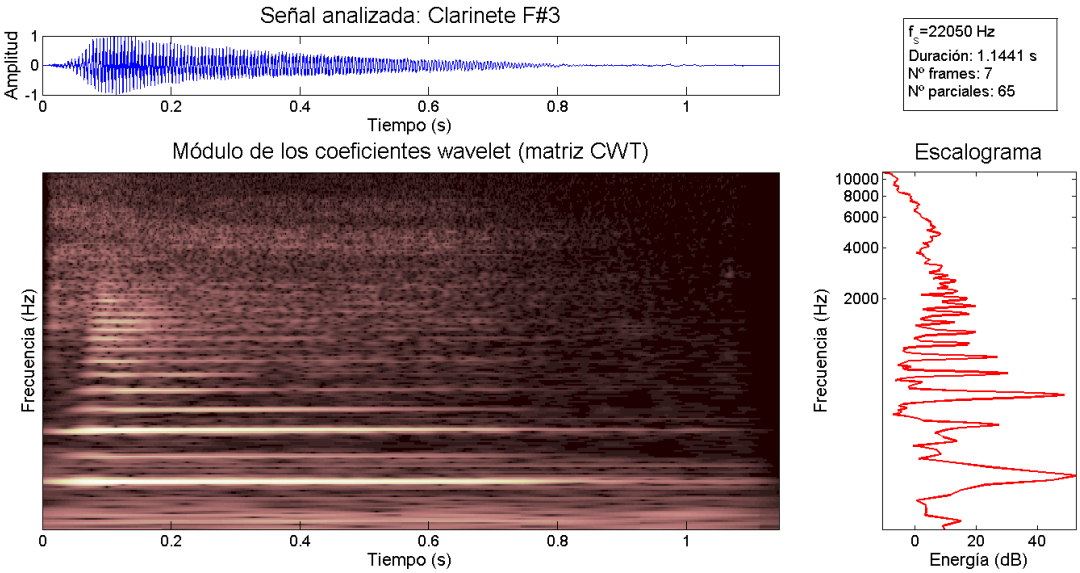


Figura I.4: *Forma de onda, módulo de los coeficientes wavelet, escalograma total y datos relevantes de una señal de clarinete interpretando una nota F#3.*

que se han analizando con el algoritmo CWAS, desde señales musicales basadas en notas puras y estables ejecutadas con diferentes tipos de instrumentos a señales de voz, pasando por mezcla de fuentes, señales marcadamente transitorias y notas con efectos de ejecución tales como vibrato o bending.

Las Figuras I.4, I.5 y I.7 son ejemplos de notas muy estables ejecutadas por diferentes instrumentos (cuerda, viento-madera). Las tres primeras señales han sido submuestreadas a $f_s = 22050$Hz, mientras que la última presenta $f_s = 44.1$kHz.

Obsérvese como en la Figura I.5 se puede observar la transición entre frames. Esto es debido al tamaño limitado en la FFT utilizada en la convolución circular para calcular los coeficientes wavelet. En la mayoría de las señales analizadas, estas discontinuidades resultan invisibles, y, en todas ellas, inaudibles.

La Figura I.6 representa un piano ejecutando 4 notas diferentes. Junto con la señal de batería analizada en la Sección 5.5.2, Figura I.10, son señales representativas de instrumentos percusivos, con transitorios muy marcados. Los transitorios se distinguen por presentar una verticalidad evidente en el espectrograma wavelet.

Como representación de una voz (cantada), la Figura en la Figura I.8 aparecen representados los resultados iniciales del análisis para una grabación de la voz de Elvis Presley. La duración total de la señal es de aproximadamente 6.1 segundos, muestreada a $f_s = 22050$Hz. Esto arroja un total de 33 frames de análisis, dentro de los cuales se han localizado y seguido

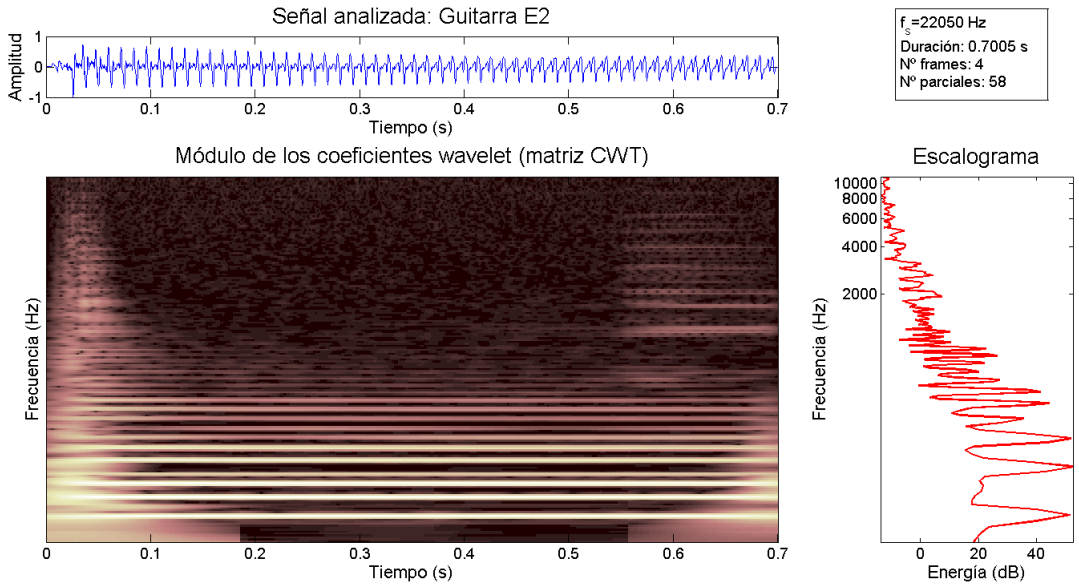


Figura I.5: *Forma de onda, módulo de los coeficientes wavelet, escalograma total y datos relevantes de una señal de guitarra interpretando una nota E2.*

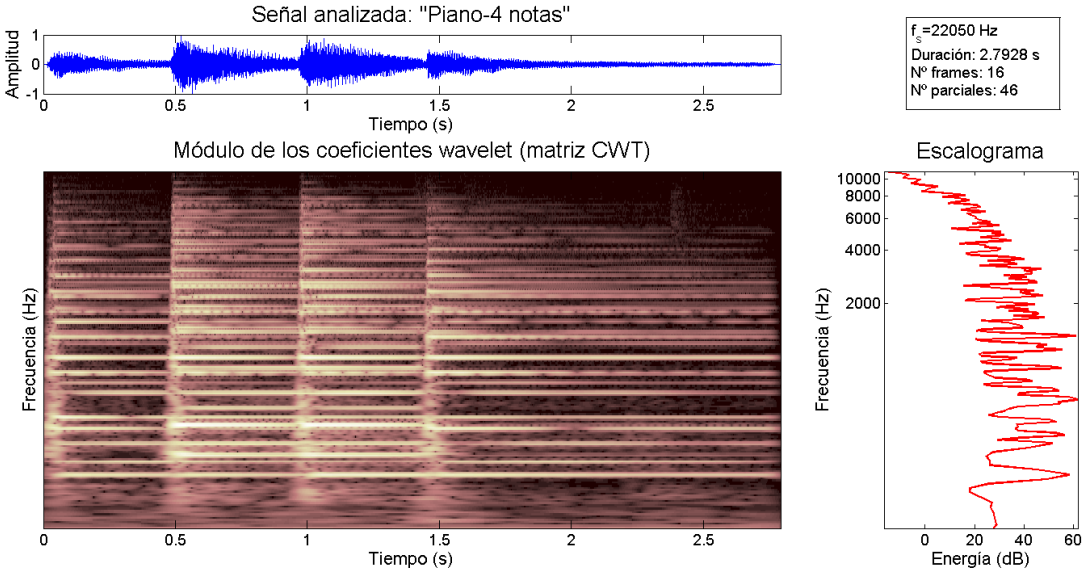


Figura I.6: *Forma de onda, módulo de los coeficientes wavelet, escalograma total y datos relevantes de la señal "Piano-4 notas".*

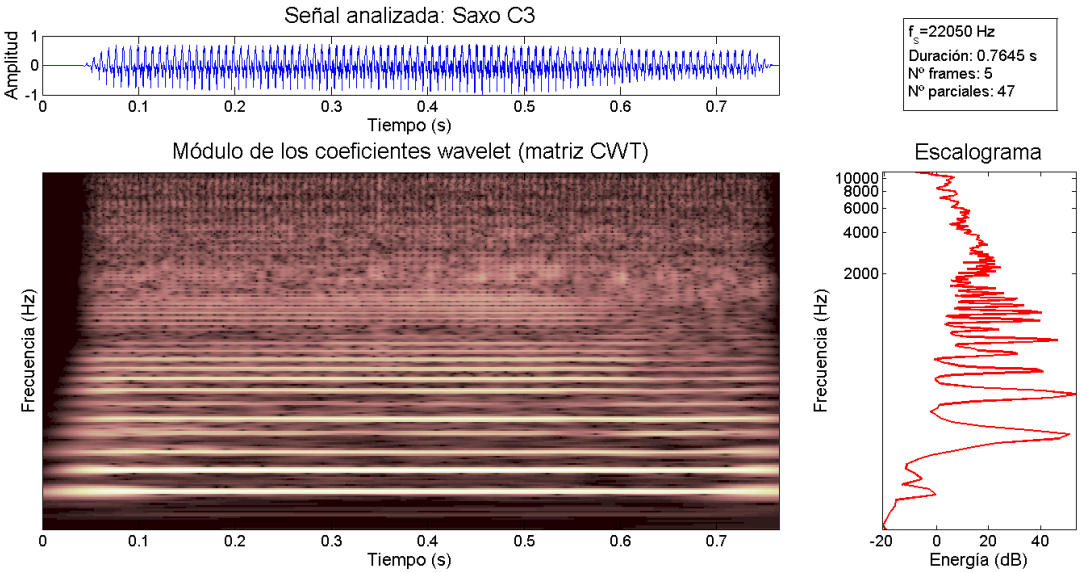


Figura I.7: *Forma de onda, módulo de los coeficientes wavelet, escalograma total y datos relevantes de una señal de un saxo lanzando una nota C3.*

hasta 60 parciales diferentes. Se puede distinguir la elevada armonía en las trayectorias de los parciales así como la característica variabilidad de la voz cantada, fácilmente distinguible de las notas estables (Figuras I.4, I.5 y I.7) pero más complicada de diferenciar de ejecuciones con bending o vibrato.

En las Figuras I.9 y I.10 se representan los resultados para las restantes dos señales cuya calidad en la resíntesis ha sido estudiada en la Sección 5.5.2, en concreto correspondientes a un solo de violín y una batería.

Por último, en la Figura I.11, el análisis de un tema musical. Concretamente un extracto de 30 segundos de la canción *"Where do you think you're going?"*, de Dire Straits[1].

En el Anexo II.c aparecen los resultados numéricos de calidad en la resíntesis de algunas de estas señales. Estos resultados se pueden constatar escuchando los sonidos relacionados, en el soporte digital adjunto.

[1] "Where do you think you're going?", *procedente del álbum* "Communiqué", *de Dire Straits* ©*Vertigo/Warner Bros., 1979.*

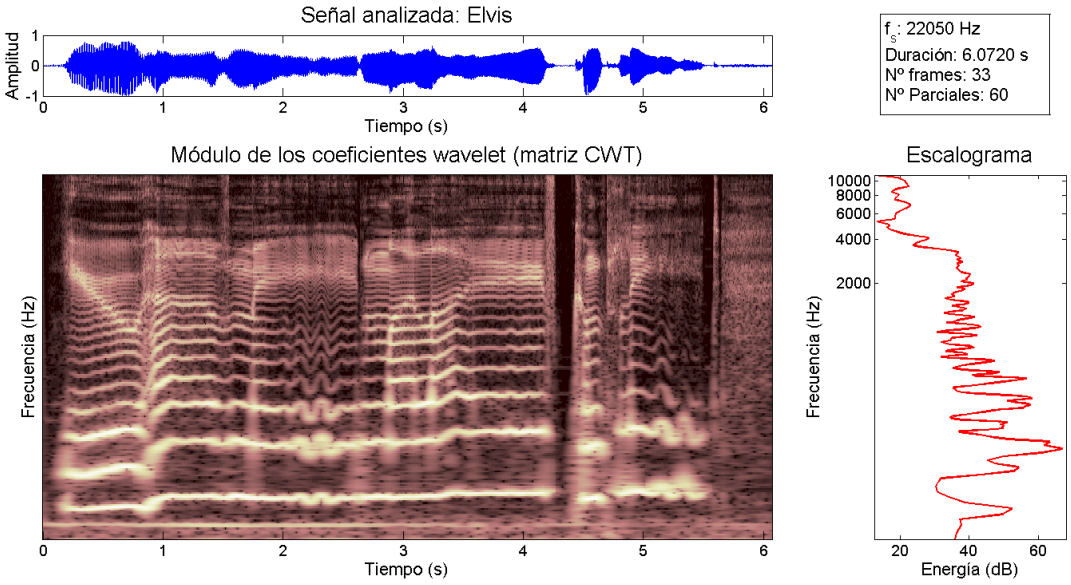


Figura I.8: *Forma de onda, módulo de los coeficientes wavelet, escalograma total y datos relevantes de la señal "Elvis".*

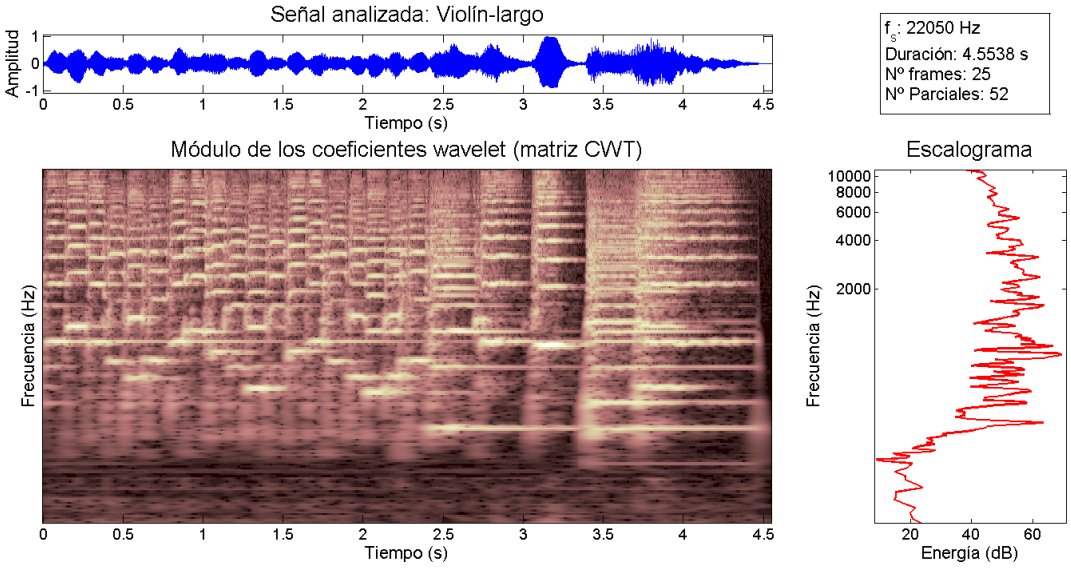


Figura I.9: *Forma de onda, módulo de los coeficientes wavelet, escalograma total y datos relevantes de la señal "Violín-largo".*

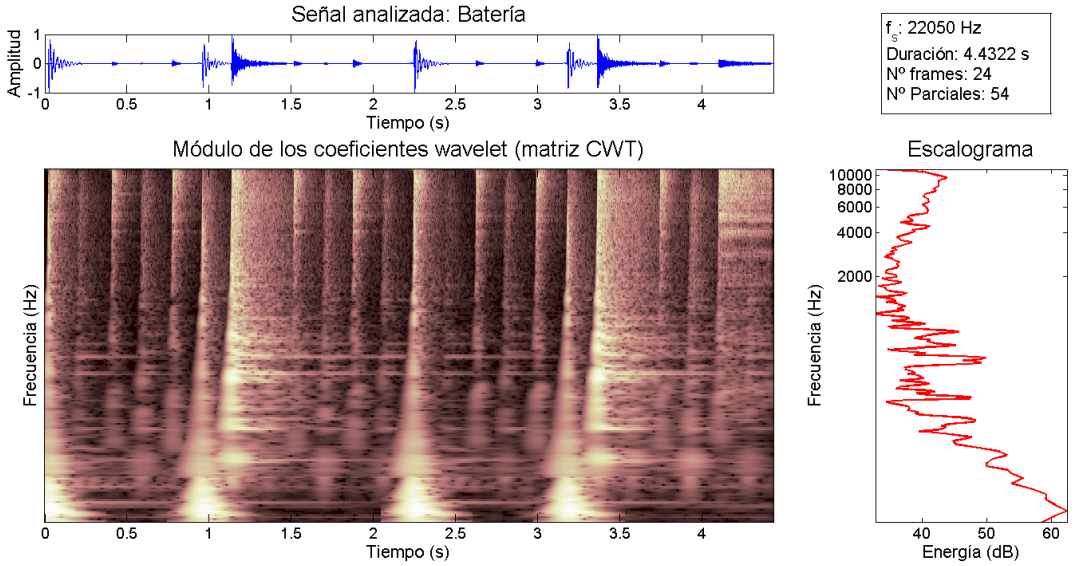


Figura I.10: *Forma de onda, módulo de los coeficientes wavelet, escalograma total y datos relevantes de la señal "Batería".*

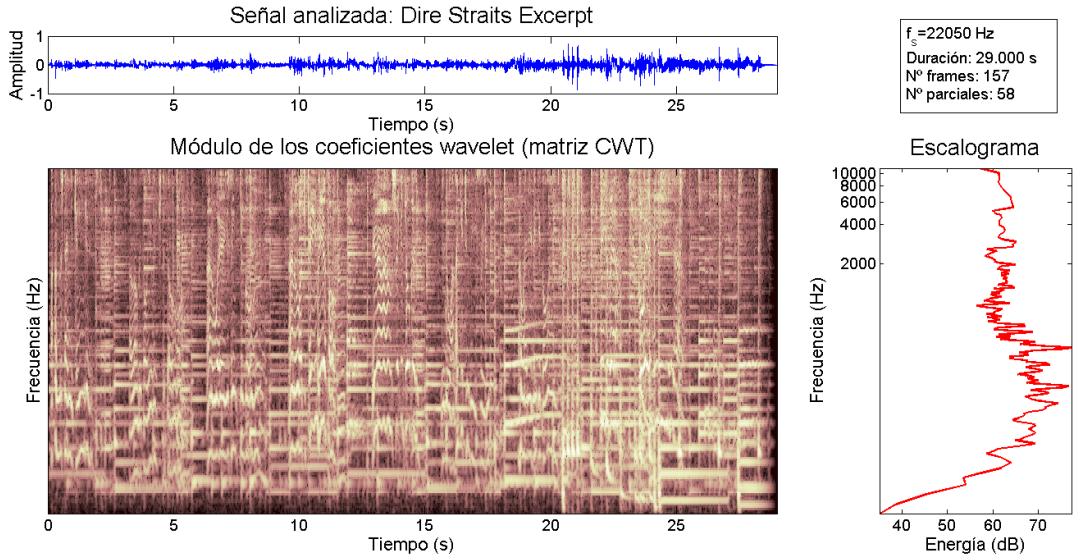


Figura I.11: *Forma de onda, módulo de los coeficientes wavelet, escalograma total y datos relevantes correspondientes al análisis de un extracto de 29 segundos del tema* "Where do you think you're going?" *de Dire Straits.*

Anexo II

Aplicaciones del algoritmo CWAS: Ampliación

"El conocimiento es la mejor inversión que se puede hacer".
Abraham Lincoln (1809-1865).
Abogado y político estadounidense.

En este Apéndice se completan algunos resultados experimentales relacionados con los Capítulos 4 y 5 de esta Tesis, en concreto los relativos a la síntesis de señales musicales y a la estimación de frecuencias fundamentales en señales multipitch. Además, se introducirán brevemente algunas pequeñas aplicaciones que se han desarrollado, relativas al análisis sub-banda, al filtrado de señales y al desarrollo de efectos de audio basados en el algoritmo CWAS.

II.a. Resultados en la detección de frecuencias fundamentales en señales monofónicas (Algoritmo #1).

El algoritmo de búsqueda de fundamentales ha sido puesto a prueba con un conjunto aleatorio de 106 señales monofónicas procedentes de la base de datos musical de la Universidad de Iowa, disponible en `http://theremin.music.uiowa.edu/MIS.html`, [63], con permiso de Lawrence Fritts, Director de Electronic Music Studios. Todas las señales analizadas son extractos de un segundo de duración, submuestreadas a f_s= 22050Hz (nuevamente

por motivos de economización temporal) y enventanadas en inicio y final por semi-ventanas de Hanning de 0.1 segundos de duración, procedentes de las grabaciones de la base de datos mencionada, llevadas a cabo en cámara anecoica. El conjunto de señales cubre un total de 11 instrumentos musicales diferentes (*flauta, oboe, clarinete en Si bemol, clarinete bajo, saxofón soprano, saxofón alto, trompa, trompeta en Si bemol, trombón bajo, violín y viola*), ejecutando una serie de notas (escalas o partes de escalas) cada uno de ellos. Los resultados se muestran en las Tablas II.1 a II.11. En estas tablas figura, para cada uno de los instrumentos, el resultado de frecuencia fundamental detectado y la nota que supuestamente se está ejecutando (nota correspondiente de la escala musical, extraída por aproximación de la Tabla I.1, véase Anexo I).

Clarinete en Sib: análisis armónico	
f_0 **detectada**	**Nota correspondiente**
1045.8	C6
1101.6	C#6
1168.8	D6
1227.8	D#6
1303.1	E6
1370.7	F6
1471.9	F#6
1554.3	G6
1641.6	G#6
1750.9	A6
1848.2	A#6
1937.8	B6

Tabla II.1: *Fundamentales del Clarinete en Si bemol.*

Clarinete bajo: análisis armónico	
f_0 **detectada**	**Nota correspondiente**
246.4652	B3
260.8564	C4
276.0825	C#4
293.2345	D4
309.0154	D#4
327.8158	E4
347.9642	F4
369.3381	F#4
390.2627	G4
414.8530	G#4
441.0428	A4
467.7150	A#4

Tabla II.2: *Fundamentales del Clarinete bajo.*

Como se ha explicado al comienzo de esta Sección, el conjunto de señales analizado se ha suavizado mediante sendos *fade-in* y *fade-out* (semiventanas de Hanning). Caso de no llevarse a cabo este suavizado, los resultados de detección son muy ligeramente distintos. Esto es debido a que en este caso, se emplean más puntos de $f_{ins,n}(t_i)$ para calcular la frecuencia promedio de los parciales a partir de la Ecuación (4.14). En la Tabla II.12 se ha representado una comparativa de resultados entre una señal de cada instrumento musical suavizada y sin suavizar. Como se puede ver, la diferencia es prácticamente despreciable.

El resultado del análisis armónico completo de una señal se puede ver en el ejemplo representado en la Figura II.1. En él aparece el escalograma de una señal de clarinete (aunque no se trata de ninguna de las señales extraídas de la base de datos de la Universidad de Iowa). Resaltada con un punto y una flecha rojos, la posición de la frecuencia fundamental.

Saxo alto: análisis armónico	
f_0 **detectada**	**Nota correspondiente**
531.3203	C5
560.7914	C#5
594.0481	D5
631.1603	D#5
672.6878	E5
710.6929	F5
753.7347	F#5
799.5817	G5
847.0195	G#5

Tabla II.3: *Fundamentales del Saxo alto.*

Saxo soprano: análisis armónico	
f_0 **detectada**	**Nota correspondiente**
263.1879	C4
277.3781	C#4
295.8007	D4
313.2937	D#4
332.3046	E4
350.8835	F4
369.5634	F#4
392.6589	G4
415.6902	G#4
435.1010	A4
466.4069	A#4
493.3131	B4

Tabla II.4: *Fundamentales del Saxo soprano.*

Trompa: análisis armónico	
f_0 **detectada**	**Nota correspondiente**
64.5523	C2
68.0391	C#2
71.9470	D2
77.3433	D#2
80.8990	E2
86.1652	F2
93.7932	F#2
98.9704	G2
104.2497	G#2
110.9698	A2
116.6604	A#2
123.5077	B2

Tabla II.5: *Fundamentales de la Trompa.*

Trombón bajo: análisis armónico	
f_0 **detectada**	**Nota correspondiente**
130.4339	C3
137.9143	C#3
145.7846	D3
155.2935	D#3
163.5675	E3
174.2231	F3
181.6807	F#3
196.0033	G3
206.9775	G#3
220.0980	A3
232.9679	A#3
248.5347	B3

Tabla II.6: *Fundamentales del Trombón bajo.*

Los primeros 20 múltiplos teóricos de ésta aparecen marcados en negro. En este ejemplo concreto, hasta el múltiplo $n^0 14$ (flecha negra), cada uno de los armónicos asociados se corresponde claramente con un pico frecuencial del escalograma. Luego, la resolución del banco de filtros ya no es capaz de detectar adecuadamente armónicos de orden superior. Esto podría suponer una limitación importante de cara a posibles usos posteriores de la información armónica, como se verá más adelante.

Trompeta en Sib: análisis armónico	
f_0 **detectada**	**Nota correspondiente**
522.7883	C5
554.5303	C#5
588.6466	D5
629.6567	D#5
663.9253	E5
700.6883	F5
749.2905	F#5
789.0573	G5
838.1145	G#5
887.3253	A5
938.9846	A#55
995.8584	B5

Tabla II.7: *Fundamentales de la Trompeta en Si bemol.*

Flauta: análisis armónico	
f_0 **detectada**	**Nota correspondiente**
247.4804	B3
260.0127	C4
275.2218	C#4
292.3434	D4
310.2709	D#4
329.3524	E4
348.7909	F4
371.1303	F#4
393.4341	G4
414.1389	G#4
442.0720	A4
469.0427	A#4
497.1489	B4

Tabla II.8: *Fundamentales de la Flauta.*

Viola: análisis armónico	
f_0 **detectada**	**Nota correspondiente**
584.8661	D5
645.2775	E5
682.6052	F5
746.3337	F#5
790.0732	G5

Tabla II.9: *Fundamentales de la Viola.*

Violín: análisis armónico	
f_0 **detectada**	**Nota correspondiente**
195.6080	G3
206.4905	G#3
218.7025	A3
228.3546	A#3
240.5399	B3

Tabla II.10: *Fundamentales del Violín.*

Oboe: análisis armónico	
f_0 **detectada**	**Nota correspondiente**
234.7147	A#3
245.5057	B3

Tabla II.11: *Fundamentales del Oboe.*

La capacidad y precisión de la técnica propuesta para detectar automáticamente la fundamental y armónicos presentes en una señal monofónica es por lo tanto bastante elevada. Esto acredita a esta pequeña aplicación del algoritmo CWAS para a ser utilizada en propósitos más ambiciosos, como se ha visto en la Sección 4.7.

Análisis armónico: comparativa				
Señal analizada	f_0 **(suavizada)**	**Nota correspondiente**	f_0 **(no suavizada)**	**Diferencia (%)**
Clarinete bajo	246.4652	B3	246.4653	-4.06E-5
Clarinete en Sib	1045.8	C6	1045.8	-1.16E-7
Flauta	247.4804	B3	247.4930	-5.09E-3
Oboe	234.7147	A#3	234.7310	-6.94E-3
Saxofón alto	531.3203	C5	531.3348	-2.73E-3
Saxofón soprano	263.1879	C4	263.1880	-3.8E-5
Trombón bajo	130.4339	C3	130.4343	-3.07E-4
Trompa	64.5523	C2	64.5987	-7.19E-2
Trompeta en Sib	522.7883	C5	522.8046	-3.12E-3
Viola	584.8661	D5	584.8517	2.46E-3
Violín	195.6080	G3	195.6084	-2.04E-4

Tabla II.12: *Comparativa de resultados de detección de fundamentales para señales suavizadas y sin suavizar.*

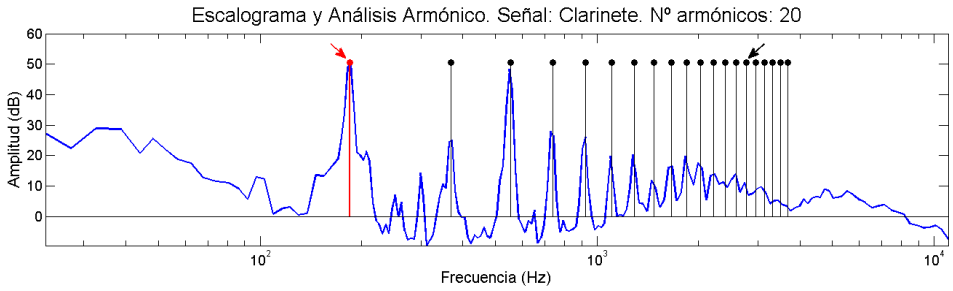
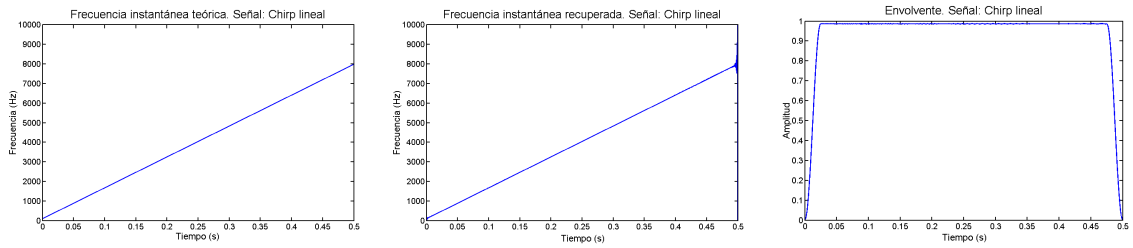


Figura II.1: *Resultados del análisis armónico completo para una señal de clarinete.*

II.b. Acerca de la precisión en el análisis tiempo–frecuencia.

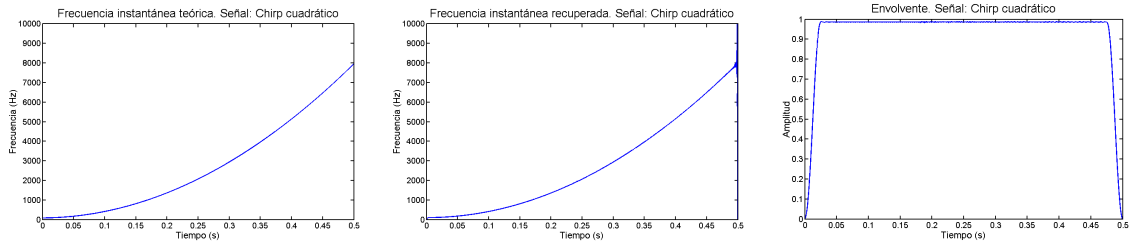
En la Sección 5.5.1 se han presentado los resultados de calidad en la resíntesis de cuatro señales sintéticas: un chirp lineal, un chirp cuadrático, un chirp exponencial y una señal de FM con excursión sinusoidal. La evolución frecuencial de estas señales se conoce perfectamente. En cuanto a la envolvente temporal se trata de un fade-in y fade-out construido con semi-ventanas de Hanning normalizado a un máximo de 0.99.

En las Figuras II.2(a) a II.5(a) se muestran tanto la evolución teórica como la experimental de las frecuencias instantáneas así como de las envolventes temporales que arroja el algoritmo CWAS para cada una de estas señales.



(a) $f_{ins}(t)$. *Valor teórico.* (b) $f_{ins}(t)$. *Resultado experimental.* (c) $A(t)$. *Resultado experimental.*

Figura II.2: *Recuperación de $f_{ins}(t)$ y $A(t)$. Señal: chirp lineal. (a) Frecuencia instantánea teórica. (b) Frecuencia instantánea recuperada. (c) Envolvente instantánea experimental.*



(a) $f_{ins}(t)$. *Valor teórico.* (b) $f_{ins}(t)$. *Resultado experimental.* (c) $A(t)$. *Resultado experimental.*

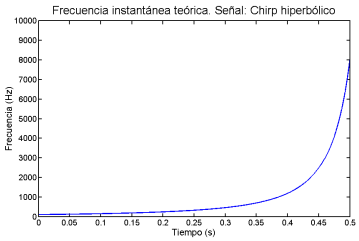
Figura II.3: *Recuperación de $f_{ins}(t)$ y $A(t)$. Señal: chirp cuadrático. (a) Frecuencia instantánea teórica. (b) Frecuencia instantánea recuperada. (c) Envolvente instantánea experimental.*

A continuación, se va a representar gráficamente el error $e_f(t)$ cometido para cada una de estas cuatro señales. Estas figuras se han obtenido restando las funciones representadas en las anteriores Figuras, II.2 a II.5, correspondientes a frecuencia instantánea teórica, parte (a) de cada figura y recuperada, parte (b). Podemos por tanto definir:

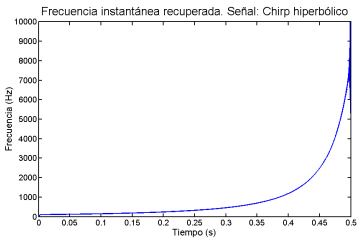
$$e_f(t) = f_{ins,x}(t) - f_{ins,x_{syn}}(t) \qquad \forall t \tag{II.1}$$

de forma paralela a como fue definido el error temporal de la señal, Ecuación (3.25).

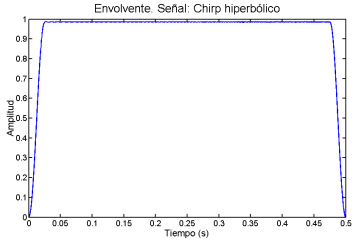
Como se puede apreciar en la Figura II.6, el error en la localización frecuencial es especialmente grande al inicio y al final de cada señal. Como se ha explicado, esto es debido al bajo valor de la amplitud instantánea $A(t)$ en estos puntos. Véanse Figuras II.2(c), II.3(c), II.4(c) y II.5(c).



(a) $f_{ins}(t)$. *Valor teórico.*

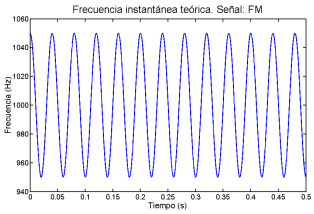


(b) $f_{ins}(t)$. *Resultado experimental.*

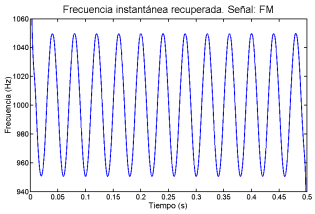


(c) $A(t)$. *Resultado experimental.*

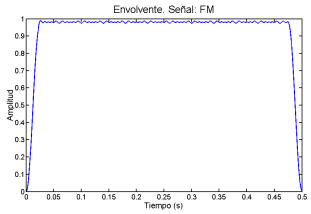
Figura II.4: *Recuperación de $f_{ins}(t)$ y $A(t)$. Señal: chirp hiperbólico. (a) Frecuencia instantánea teórica. (b) Frecuencia instantánea recuperada. (c) Envolvente instantánea experimental.*



(a) $f_{ins}(t)$. *Valor teórico.*



(b) $f_{ins}(t)$. *Resultado experimental.*



(c) $A(t)$. *Resultado experimental.*

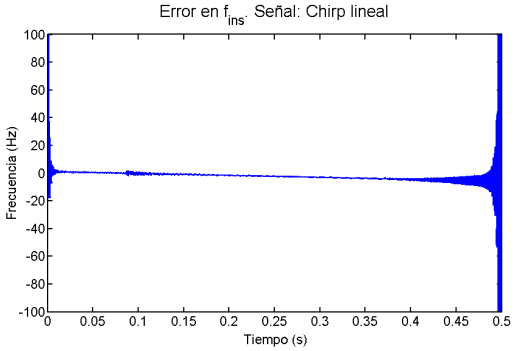
Figura II.5: *Recuperación de $f_{ins}(t)$ y $A(t)$. Señal: FM con excursión sinusoidal. (a) Frecuencia instantánea teórica. (b) Frecuencia instantánea recuperada. (c) Envolvente instantánea experimental.*

II.c. Más acerca de síntesis de señales

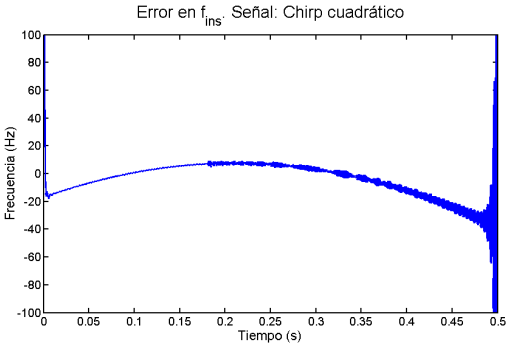
A lo largo de esta disertación, se ha afirmado varias veces que los resultados de resíntesis de la señal de audio que arroja el algoritmo CWAS son de altísima calidad. Para demostrar parcialmente ésta afirmación (otras demostraciones tanto gráficas como numéricas pueden encontrarse en el Capítulo 5) se van a presentar resultados de error instantáneo $\varepsilon(t)$ obtenidos para un conjunto de quince señales, que abarcan desde notas ejecutadas por instrumentos musicales aislados a extractos de temas musicales, pasando por voz.

Los resultados de seis de estas señales (tomadas al azar del grupo de análisis) aparecen representados gráficamente en la Figura II.7.

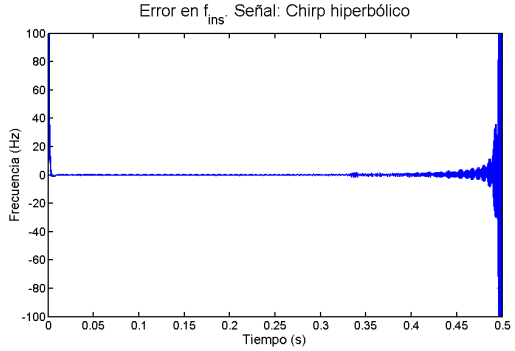
Como se puede apreciar, la diferencia entre la señal original y la sintética se encuentra en cualquier caso por debajo de los 27dB, y en promedio queda por debajo de 50dB. Estas



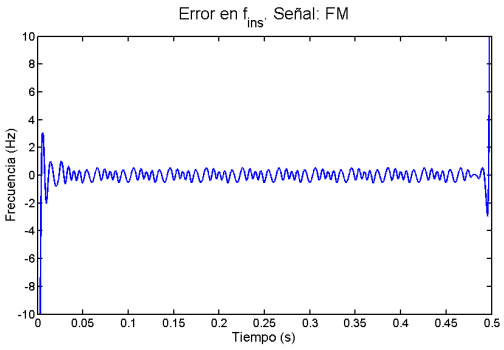
(a) *Error en frecuencia. Señal: Chirp lineal.*



(b) *Error en frecuencia. Señal: Chirp cuadrático.*



(c) *Error en frecuencia. Señal: Chirp hiperbólico.*



(d) *Error en frecuencia. Señal: FM.*

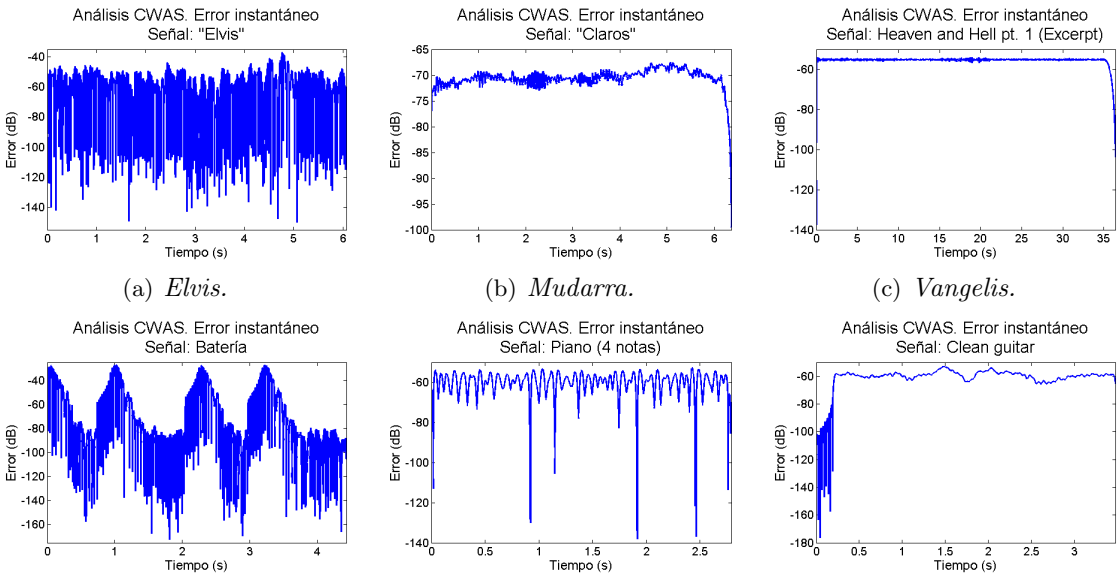
Figura II.6: *Gráficas de error frecuencial obtenidas en el análisis de las señales sintéticas: (a) Chirp lineal. (b) Chirp cuadrático. (c) Chirp hiperbólico. (d) FM.*

diferencias resultan generalmente inapreciables para el oído no entrenado.

En la Tabla II.13 se reflejan los resultados numéricos asociados a éste análisis. En la primera columna aparecen las señales analizadas, en la segunda los errores medios experimentales para cada señal ($\overline{\varepsilon}$), y en la tercera el error máximo cometido en cada caso (ε_{max}).

II.d. Otras aplicaciones del algoritmo CWAS

Por último, se procederá a introducir una batería de pequeñas aplicaciones del algoritmo CWAS que se han desarrollado. Como se ha avanzado anteriormente, se trata de ensayos relativos al análisis sub-banda y al filtrado de señales, así como algunos de los 17 efectos de sonido obtenidos aplicando la técnica propuesta.



(a) *Elvis.* (b) *Mudarra.* (c) *Vangelis.*

(d) *Batería.* (e) *Piano.* (f) *Guitarra.*

Figura II.7: *Resultados del análisis CWAS. Error instantáneo (dB): (a) Señal analizada: Elvis Presley ("you have made my life complete"). (b) Extracto de la pieza clásica* "Claros y frescos ríos"*, Alonso de Mudarra, 1546. (c) Extracto de 36 segundos del tema de Vangelis "Heaven and Hell, part 1", 1975. (d) Ritmo de batería. Contiene caja, bombo y charles. (e) Cuatro notas de piano. (f) Seis notas de guitarra.*

II.d.1. Análisis sub–banda

La audición humana está condicionada por las bandas críticas del oído. Las bandas críticas controlan la consonancia o disonancia de los sonidos, así como el efecto de enmascaramiento (ampliamente utilizado en codificación y compresión de audio), entre otros efectos. El ancho de banda de las bandas críticas se sitúa entre un tercio y un sexto de octava, [13, 56], lo cual da origen a las bien conocidas 24 bandas críticas del oído. Estas bandas críticas también afectan a la percepción de batidos cuando dos tonos puros están lo bastante próximos en frecuencia. Cuando dos de tales tonos se encuentran lo bastante cerca uno del otro, el oído es incapaz de resolverlos en términos espectrales y el sonido percibido puede ser mejor descrito por una señal que exhibe variaciones tanto en amplitud como en frecuencia (batidos). El par canónico de tal señal debería modelar este hecho.

En la Sección 1.2 se obtuvieron una serie de expresiones recursivas sobre el par canónico de una señal conteniendo N parciales, con sus respectivos batidos en frecuencia. Es posible simplificar tales expresiones para el caso particular de señales con $N = 2$ parciales batiendo, [24]. En esta Sección se pretende demostrar cómo resulta factible extraer información por

Señal	$\overline{\varepsilon}$ (dB)	ε_{max} (dB)
"Claros"	-70.5106	-67.5365
Batería	-77.0985	-27.2011
Vangelis	-55.1812	-54.0277
Saxo C4	-62.9367	-62.3186
Violín largo	-64.7637	-38.8218
Guitarra limpia	-59.3913	-53.0850
Cuerno C3	-67.5835	-64.5479
Piano	-59.2471	-52.6455
Flauta F#5	-79.7811	-68.6359
Fagot F#5	-73.5123	-61.7532
Clarinete F#3	-56.3659	-44.4506
Guitarra B4	-60.0370	-56.5365
"Elvis"	-57.1500	-37.1431
Viola D5	-82.8153	-75.6732
Dire Straits	-79.6416	-50.5206

Tabla II.13: *Errores promedio y máximo cometidos en la resíntesis de señales de audio no sintéticas (en dB).*

debajo del límite impuesto por el análisis en bandas a partir del par canónico de $x(t)$, es decir, superar la barrera de las bandas críticas.

II.d.1.1. Expresiones canónicas del batido de componentes

La expresión de partida para la señal es:

$$x(t) = x_1(t) + x_2(t) = A_1(t)\cos[\phi_1(t)] + A_2(t)\cos[\phi_2(t)] \tag{II.2}$$

La propia $x(t)$ puede calcularse a partir de la parte real de su *señal analítica*, la cual depende a su vez de las señales analíticas relativas a cada uno de sus componentes (para más detalles, véase la Sección 1.2), es decir:

$$x(t) = \Re[x_{an}(t)] = \Re\left[A_1(t)e^{j\phi_1(t)} + A_2(t)e^{j\phi_2(t)}\right] = \Re[x_{an_1}(t) + x_{an_2}(t)] \tag{II.3}$$

Por lo tanto:

$$x_{an}(t) = x_{an_1}(t) + x_{an_2}(t) = \left[A_1 + A_2e^{j\Delta(t)}\right]e^{j\phi_1(t)} \tag{II.4}$$

De este modo, la Ecuación (II.3) queda:

$$x(t) = \|x(t)\| cos\left\{\phi_1(t) + \Phi(t)\right\} \tag{II.5}$$

donde $\Delta(t) = \phi_2(t) - \phi_1(t)$, y siendo:

$$\Phi(t) = \arctan\left\{\frac{A_2(t)\sin\left[\Delta(t)\right]}{A_1(t) + A_2(t)\cos\left[\Delta(t)\right]}\right\} \tag{II.6}$$

Por lo tanto, el módulo de $x(t)$ puede escribirse como:

$$\|x(t)\| = \sqrt{A_1^2(t) + A_2^2(t) + 2A_1(t)A_2(t)\cos\left[\Delta(t)\right]} \tag{II.7}$$

Aplicando la Ecuación (1.9), la frecuencia instantánea de la señal queda:

$$\begin{aligned} f_{ins}(t) &= \frac{1}{2\pi}\frac{d\left\{\phi_1(t) + \Phi\left[x(t)\right]\right\}}{dt} = \\ &= \frac{1}{2\pi}\left\{\phi_1'(t) + \frac{\Delta'(t)\left\{A_2^2(t) + A_1A_2\cos[\Delta(t)]\right\}}{A_1^2(t) + A_2^2(t) + 2A_1(t)A_2(t)\cos[\Delta(t)]}\right\} \end{aligned} \tag{II.8}$$

Las expresiones hasta aquí obtenidas para $\|x(t)\|$, $\Phi(t)$ y $\Delta(t)$ son la aplicación de las Ecuaciones (1.14) a (1.16) al caso particular que se está estudiando. Para mayor número de componentes batiendo, estas fórmulas pueden aplicarse, como se explicó en el Capítulo 1, de forma recursiva.

Las Ecuaciones (II.7) a (II.8) indican que envolvente y fase del par canónico de $x(t)$ presentan una dependencia temporal que refleja la modulación de la onda. Si la separación entre las componentes de la señal $\phi_1(t)$ y $\phi_2(t)$ es suficientemente grande, la distinción entre moduladora y portadora se hace más evidente. Sin embargo, expresar la mezcla de las componentes de la señal en una representación modulo-fase como la aquí obtenida, tiene tanto más sentido cuanto más próximos estén tales componentes. Es decir, pese a que las Ecuaciones (II.7) y (II.6) son perfectamente válidas en el caso general, resultan más útiles para el caso de parciales batientes, cuando los términos de intermodulación resultan audibles, por suponer una representación matemática aproximada del comportamiento del oído humano.

II.d.1.2. Ejemplos teóricos de análisis sub–banda

Se va a poner a prueba la validez del algoritmo CWAS a la hora de obtener el par canónico de una señal, para el caso de sendas ondas compuestas por dos cosenos.

En primer lugar, dos cosenos de frecuencias 440Hz y 493.76Hz, es decir, las notas $A4$ y $B4$. La representación gráfica de la forma de onda obtenida se encuentra en la parte superior

de la Figura II.8. En la parte inferior de la misma figura se pueden ver los batidos correspondientes a sendas componentes separadas por sólo 15Hz, concretamente de frecuencias 440Hz y 455Hz.

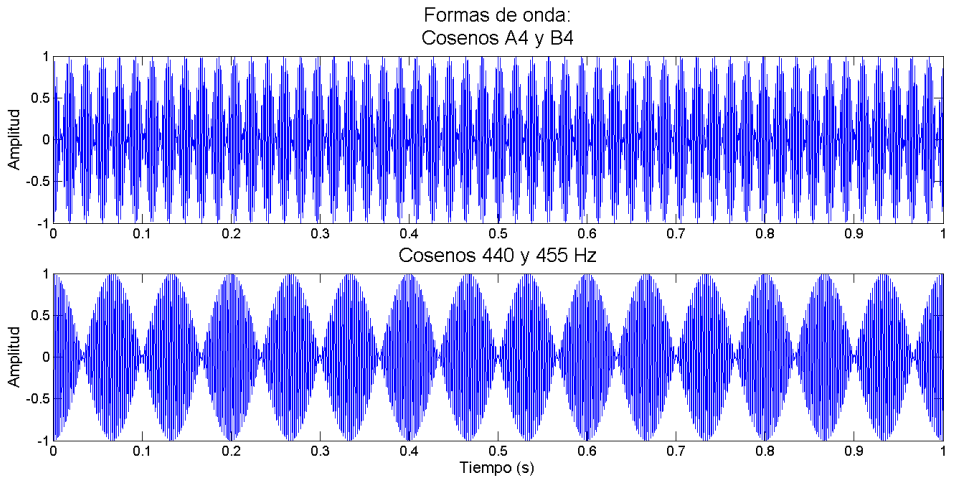


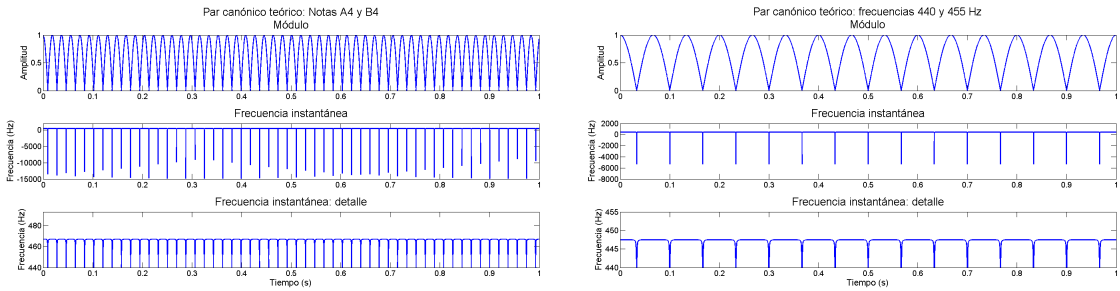
Figura II.8: *Formas de onda teóricas.*

En la Figura II.9(a) aparecen representados los resultados de las Ecuaciones (II.7) y (II.8) para el caso de los tonos $A4$ y $B4$. La gráfica superior de la figura representa el módulo, obtenido directamente a través de la Ecuación (II.7), con $A_1 = 0.5$ y $A_2 = 0.499$, $f_1 = 440$Hz y $f_2 = 493.76$Hz. En las dos gráficas inferiores aparecen la frecuencia instantánea, Ecuación (II.8), y un detalle de la misma. En la gráfica inferior se puede observar que la frecuencia detectada es aproximadamente el promedio de f_1 y f_2, y que presenta una serie de saltos en su valor muy evidentes allí donde la amplitud del módulo está próxima a cero. Tales irregularidades tienen un período relacionado con el valor de $\Delta(t)$, como se deduce de las expresiones matemáticas.

En la Figura II.9(b) se han representado gráficamente los resultados de módulo y fase para la señal compuesta por dos tonos separados por 15 Hz. En este caso, $A_1 = 0.5$, $A_2 = 0.499$, $f_1 = 440$Hz y $f_2 = 455$Hz.

II.d.1.3. Resultados prácticos

Se ha analizado con el algoritmo CWAS un conjunto de señales sintéticas que incluye ondas similares a los dos ejemplos representados en el apartado anterior. Los resultados completos del análisis se ofrecen en el Anexo II. En esta Sección se detallarán los resultados correspondientes a las dos señales similares a las que ya se han visto, concretamente, dos



(a) *Módulo, frecuencia instantánea y detalle de la misma. Señal: suma de cosenos de notas A4 y B4.* (b) *Módulo, frecuencia instantánea y detalle de la misma. Señal: suma de cosenos de 440 (A4) y 455 Hz.*

Figura II.9: *Ejemplos de batido de parciales próximos. Par canónico de una señal compuesta por: (a) Dos componentes separadas por un tono. (b) Dos componentes separadas por 15 Hz.*

componentes de amplitudes $A_1 = 0.47025$ y $A_2 = 0.495$, separadas por un tono completo (frecuencias de 440Hz, $A4$ y 493.76Hz, $B4$) y dos componentes de las mismas amplitudes que el caso anterior, esta vez separadas por 15Hz (frecuencias de 440Hz y 455Hz). Todas las señales tienen una duración de 1 segundo, han sido muestreadas bajo f_s =44.1kHz, están normalizadas a 1 y suavizadas en los extremos con semi-ventanas de Hanning para reducir los efectos de borde. La representación gráfica de las formas de onda se muestra en la Figura II.10. Comparándolas con las de la Figura II.8, los paralelismos resultan evidentes.

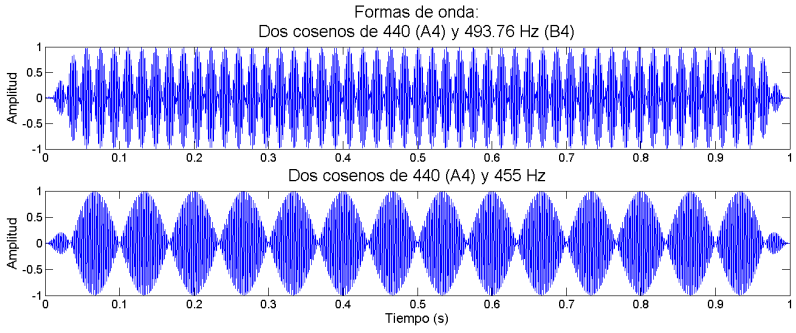
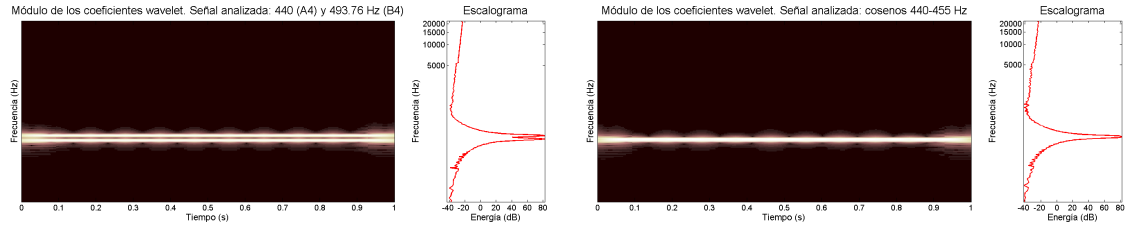


Figura II.10: *Formas de onda de las señales analizadas.*

En la Figura II.11 se presentan el módulo de los coeficientes wavelet y el escalograma de estas dos señales. Como se puede observar, en el primer caso, Figura II.11(a), el banco de filtros tiene resolución suficiente como para encontrar y separar ambos tonos, mientras que

en el segundo caso, Figura II.11(b), la resolución es insuficiente y se detecta un solo parcial.



(a) *Módulo de los coeficientes wavelet y escalograma. Señal: suma de cosenos de notas A4 y B4.*

(b) *Módulo de los coeficientes wavelet y escalograma. Señal: suma de cosenos de frecuencias 440 y 455 Hz.*

Figura II.11: *Resultados del análisis: (a) Señal: Componentes separados por un tono. (b) Señal: Componentes separados por 15 Hz.*

Dada la naturaleza de la información que arroja el algoritmo CWAS, para la primera señal se obtendrá por lo tanto la información de amplitud y fase (frecuencia) instantáneas de cada uno de los parciales detectados. La proximidad de los mismos causará, como se verá más adelante, pequeñas oscilaciones que sobre todo afectan a la información de amplitud, restos de la intermodulación presente en la señal. Por otro lado, en el segundo caso se detecta un solo pico en el escalograma, y por lo tanto un único parcial, como se ha adelantado. Este caso resulta más interesante y conviene explicarlo más en detalle. Para empezar, es evidente que en esta situación se tiene que $\Delta(t) = 2\pi(f_2 - f_1)t = 2\pi\Delta t$ radianes/segundo. Por lo tanto, las expresiones teóricas se convierten, para este caso particular, en:

$$\|x(t)\| = \sqrt{A_1^2 + A_2^2 + 2A_1A_2\cos(2\pi\Delta t)} \tag{II.9}$$

y:

$$f_{ins}(t) = f_1 + \frac{\Delta\left[A_2^2 + A_1A_2\cos(2\pi\Delta t)\right]}{A_1^2 + A_2^2 + 2A_1A_2\cos(2\pi\Delta t)} \qquad \epsilon\left[f_1 + \frac{\Delta A_2}{A_1 + A_2}, f_1 + \frac{\Delta A_2}{|A_1 - A_2|}\right] \tag{II.10}$$

En la Figura II.12 se ha representado la envolvente del parcial (obtenida a través del módulo de los coeficientes wavelet) y la frecuencia instantánea (obtenida a través la derivada de la fase instantánea desenrollada).

Las similitudes respecto a la Figura II.9(b) son obvias, de lo se colige que el resultado del análisis está próximo al par canónico teórico de la señal. La diferencia en el signo de los saltos en la fase es debida a la relación entre las amplitudes A_1 y A_2 en ambos casos. Concretamente, en el ejemplo teórico era $A_1 > A2$ mientras que en este caso se da $A_1 < A_2$.

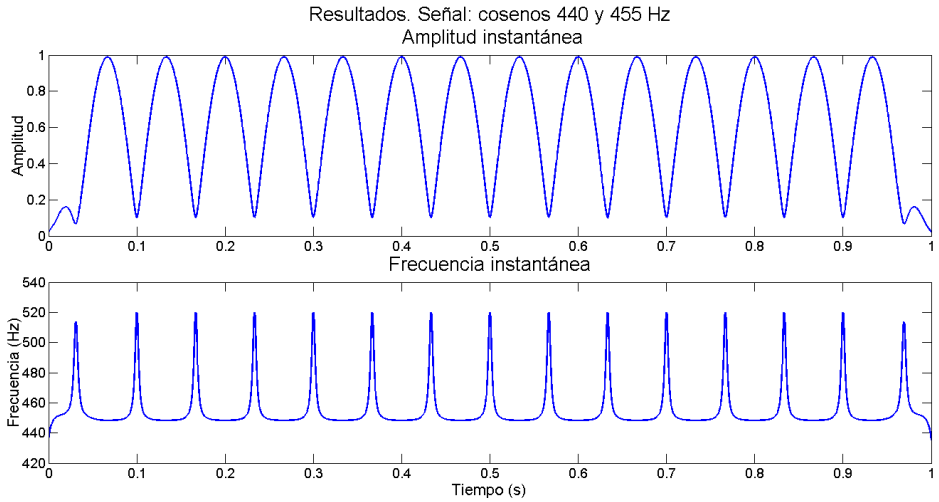


Figura II.12: *Amplitud y frecuencia instantáneas experimentales. Señal: Coseno 440 y 455 Hz.*

De esta forma se demuestra que se puede saber cuál de las dos amplitudes instantáneas es mayor que la otra sin más que ver el signo que presentan las irregularidades en f_{ins}. De hecho, es posible obtener las frecuencias y amplitudes correspondientes a cada parcial sin más que postprocesar la información ofrecida por el algoritmo.

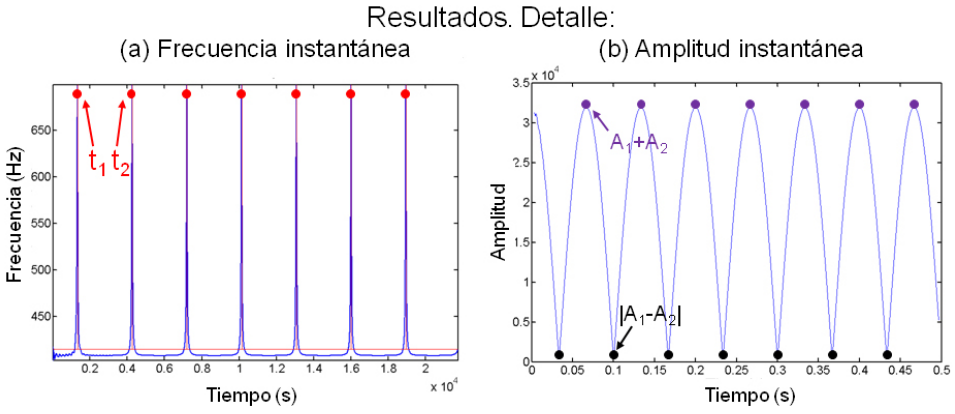


Figura II.13: *Resultados experimentales: detalle. (a) Frecuencia instantánea (b) Amplitud instantánea.*

En la Figura II.13 se ha representado un detalle de la frecuencia y la amplitud instantáneas que se han obtenido analizando la segunda de las señales introducida anterior-

mente, es decir, de amplitudes $A_1 = 0.47025$ y $A_2 = 0.495$, y de frecuencias $f_1 = 440$ Hz y $f_2 = 455$ Hz. En este caso sencillo, utilizando dos puntos cualesquiera donde la frecuencia instantánea presenta un pico, t_1 y t_2 marcados en rojo en la Figura II.13 (a), se puede calcular con facilidad el valor de Δ (en Hz):

$$\Delta = \frac{1}{t_2 - t_1} \tag{II.11}$$

De la información de amplitud, Figura II.13 (b), se pueden asimismo obtener las amplitudes correspondientes a cada parcial. En la Ecuación (II.9), es evidente que la amplitud del par canónico oscila entre $A_1 + A_2$ (en los puntos donde $\cos[2\pi\Delta t] = 1$) y $|A_1 - A_2|$ (allá donde $\cos[2\pi\Delta t] = -1$). Localizando un máximo del módulo (del conjunto de puntos marcados en morado en la figura) y un mínimo (marcados en negro), y sabiendo por el signo de los saltos frecuenciales cuál de entre A_1 y A_2 es mayor, es posible despejar sin ningún problema las amplitudes correspondientes.

En cuanto a los valores de frecuencia, el procedimiento a partir del cual trabajaremos queda reflejado en la Figura II.14.

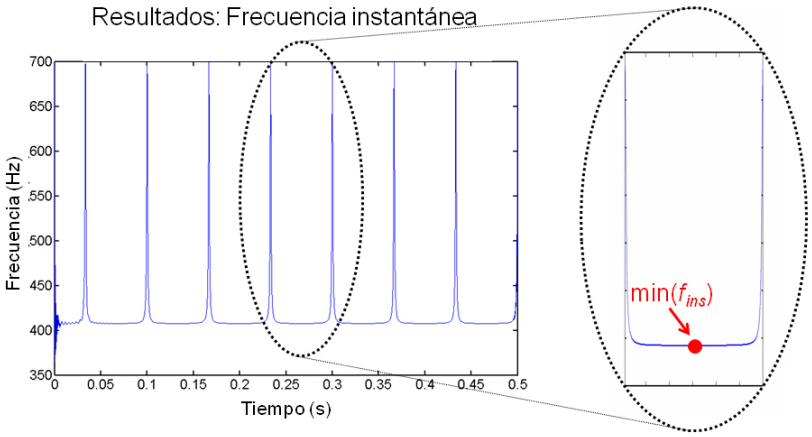


Figura II.14: *Resultados experimentales: nuevo detalle de la frecuencia instantánea, con ampliación. El dato experimental $min(f_{ins})$ lleva implícita la información frecuencial.*

Atendiendo a la Ecuación (II.10), es evidente que el mínimo en la frecuencia instantánea (señalado en rojo en la figura) se corresponde con $f_1 + \Delta A_2/(A_1 + A_2)$. Dado que tanto las amplitudes como el desfase son ahora conocidas, es posible extraer f_1 y, con este valor, f_2.

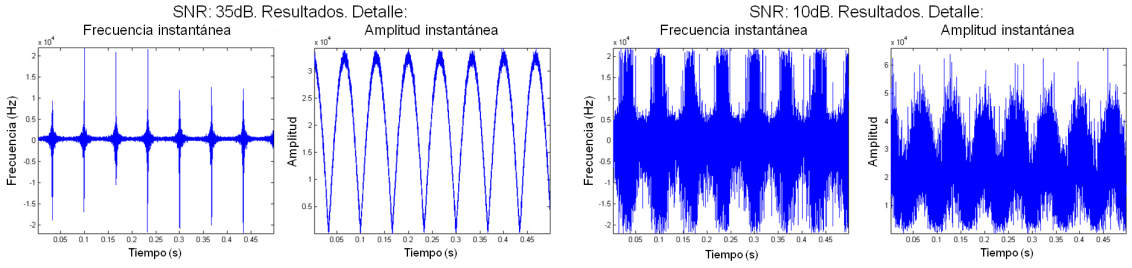
II.d.1.3.1. Estabilidad frente a la corrupción

Se ha corrompido la señal con diferentes niveles de ruido blanco, poniendo a prueba la solidez de la técnica. Los resultados se muestran en la Tabla II.14.

SNR	A_1	**e**(A_1)**(dB)**	A_2	**e**(A_2)**(dB)**	f_1**(Hz)**	**e**(f_1) **(%)**	f_2**(Hz)**	**e**(f_2)**(%)**
Original	**0.47025**	**-**	**0.495**	**-**	**400**	**-**	**415**	**-**
∞	0.4749	-40.1	0.5000	-39.9	400.3806	0.09515	415.3823	0.0948
75 dB	0.4845	-30.37	0.5100	-30.39	400.3916	0.0979	415.3941	0.0950
65 dB	0.4798	-33.84	0.5050	-33.91	400.3894	0.0973	415.3911	0.0942
55 dB	0.4795	-34.16	0.5035	-35.34	400.3610	0.0903	415.3636	0.0876
45 dB	0.5216	-19.24	0.5242	-24.58	400.2356	0.0589	415.2719	0.0655
25 dB	0.5806	-12.59	0.5830	-15	400.1350	0.0337	415.0517	0.0125
15 dB	0.7305	-5.14	0.7364	-6.24	397.8745	-0.5314	412.9796	-0.4868
10 dB	1.0079	1.16	1.0089	0.33	396.5816	-0.8546	412.0272	-0.7163

Tabla II.14: *Resultados empíricos para A_1, A_2, f_1 y f_2 con sus respectivos errores para diferentes niveles de ruido blanco ambiental.*

Como se deduce de los datos, la obtención de frecuencias instantáneas resulta bastante precisa incluso en condiciones ambientales extremas. Esto es en buena parte debido a que el ruido blanco aleatorio tiende a afectar por igual tanto a A_1 como a A_2, y en la Ecuación (II.10) éstas intervienen básicamente como cociente. Los resultados para las amplitudes crecen con el ruido y son significativamente peores a partir de los 25dB, algo comprensible cuando la información de la que se extrae llega a estar alterada por un nivel de ruido del orden de la propia señal, y más teniendo en cuenta que se calculan a través de la Ecuación (II.9).



(a) *Amplitud y frecuencia instantáneas. SNR: 35 dB.* (b) *Amplitud y frecuencia instantáneas. SNR: 10 dB.*

Figura II.15: *Resultados experimentales. Detalle: (a) SNR: 35 dB. (b) SNR: 10 dB.*

Como muestra del nivel de corrupción en los datos de partida, en la Figura II.15 se han presentado los resultados de frecuencia instantánea y envolvente del par canónico para los niveles de ruido de 35dB y 10dB. La exactitud en los datos de salida obtenidos (véase la Tabla II.14) bajo estas condiciones extremas, demuestra la validez de la técnica desarrollada.

II.d.1.4. Conclusiones y limitaciones

Como se deduce a partir de la comparación de resultados teóricos y experimentales, al menos para el caso de dos componentes, el algoritmo CWAS ofrece como salida una expresión muy próxima al par canónico de la señal analizada. Por lo tanto, deduciendo las variables adecuadas, es posible inferir cuando menos información parcial acerca de los componentes que están batiendo, superando incluso el límite de resolución impuesto por las bandas de análisis. No obstante, cabe preguntarse hasta qué punto esta información es necesaria. Teniendo en cuenta que en la mayoría de las aplicaciones se busca una resíntesis suficientemente buena, parece que el modelo subyacente al análisis resulta adecuado *per se*.

Por otro lado, estos resultados no se han puesto a prueba con señales no sintéticas, si bien la extracción de información sub–banda puede resultar interesante en, por ejemplo, algoritmos de separación de fuentes, concretamente de fuentes con espectros superpuestos (aunque el análisis presentado resulta aplicable únicamente a dos parciales batientes y sería necesario generalizar la técnica para casos más complejos, comenzando por amplitudes y frecuencias variables en cierto entorno). Desgraciadamente, a medida que el número de parciales batientes aumenta, la complejidad de las ecuaciones sub–banda crece exponencialmente, de modo que extraer información sobre todo información "ciega") procedente de tales parciales probablemente resulte prohibitivo a nivel algorítmico. Una estimación grosera del límite práctico para el número de parciales separables por este método bien podría ser $N = 3$.

II.d.2. Filtrado de señales

Una de las ventajas más evidentes de un eventual proceso de tracking de parciales punto por punto consiste en que para cada muestra de la señal es posible conocer con precisión las bandas límite superior e inferior que definen cada parcial. La variabilidad suele ser lo suficientemente elevada como para que tales parciales exhiban una marcada tendencia a presentar artefactos lo cual, unido al alto coste computacional del proceso, hace inicialmente desaconsejable su uso, como se ha concluido anteriormente.

Sin embargo, existe un proceso en el que tales espurios tienen mucha menor importancia, y es en el filtrado de señales. En efecto, cuando se corrompe una señal con ruido (por ejemplo ruido blanco), cuanto más se afine en la localización de las bandas asociadas a cada parcial, menor cantidad de ruido arrastrará la señal de salida.

II.d.2.1. Filtrado de señales monocomponente

Se ha puesto a prueba la capacidad de filtrado del algoritmo CWAS en un entorno controlado. Para ello, se han empleado cuatro señales sintéticas monocomponente: Un tono de frecuencia pura $x_1(t)$, una señal de FM típica $x_2(t)$, un chirp lineal $x_3(t)$ y un chirp hiperbólico $x_4(t)$. Todas ellas han sido corrompidas posteriormente con un nivel creciente de ruido.

Las señales sintéticas originales pueden escribirse como sigue:

$$x_1(t) = A(t)\cos(2\pi f_1 t) \tag{II.12}$$

$$x_2(t) = A(t)\cos[2\pi f_2 t + 2\sin(2\pi f_3 t)] \tag{II.13}$$

$$x_3(t) = A(t)\cos[2\pi(at^2 + bt)] \tag{II.14}$$

y:

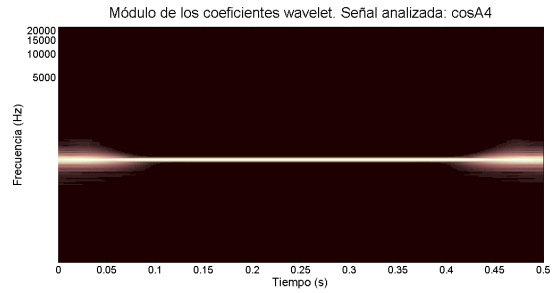
$$x_4(t) = A(t)\cos(2\pi\frac{\alpha}{\beta - t}) \tag{II.15}$$

Donde $A(t)$ es una envolvente común a todas las señales, una vez más de valor máximo 0.99 y suavizada en inicio y final por sendas semi-ventanas de Hanning. Estas señales se corresponden exactamente con el tono de 440Hz, el chirp lineal, el chirp hiperbólico y la señal de FM analizadas en el Capítulo 5, filas 1, 3, 5 y 2 de la Tabla 5.2, respectivamente.

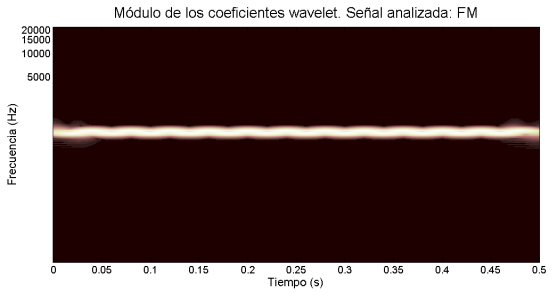
A continuación, las señales originales así como sus versiones con ruido blanco aleatorio añadido han sido analizadas por el algoritmo CWAS bajo un número de divisiones por octava constante, $D = 24$, efectuándose un seguimiento de parciales punto por punto. El espectrograma de las cuatro señales (sin ruido) se muestra en la Figura II.16. Puesto que en cada instante de tiempo el corte en bandas proporciona exclusivamente la información contenida entre las frecuencias límite del parcial (aproximadamente la zona de colores claros en las gráficas), rechazando todas las demás, la señal de salida tiende a presentar un nivel de ruido sensiblemente inferior a la analizada en cada caso.

Una vez obtenido el conjunto de señales sintéticas extraídas en cada caso (bajo diferentes condiciones de ruido ambiente), es posible compararlas con la señal aislada.

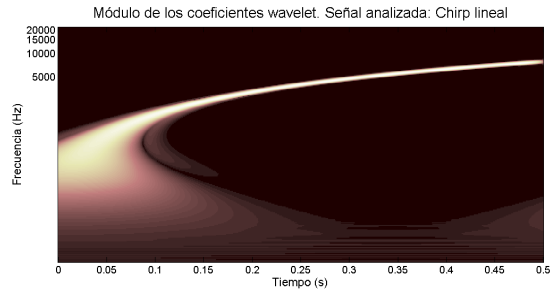
Aunque se asume que los resultados no serán en este caso tan buenos numérica y gráficamente como en ejemplos anteriores, es posible una vez más, obtener la forma de onda del error en tiempo a través de la Ecuación (3.25), el valor máximo del mismo utilizando la Ecuación (5.9), y comparar los espectros de las señales, obteniendo un error máximo equivalente en la información frecuencial. Los resultados se muestran en la última gráfica de las Figuras II.17 a II.19. Pese a que, numéricamente, se llegan a alcanzar errores considerables (datos a la izquierda de las gráficas), la calidad del filtrado es elevada (en términos acústicos). En estas figuras se puede apreciar además como, lógicamente, el error en la cap-



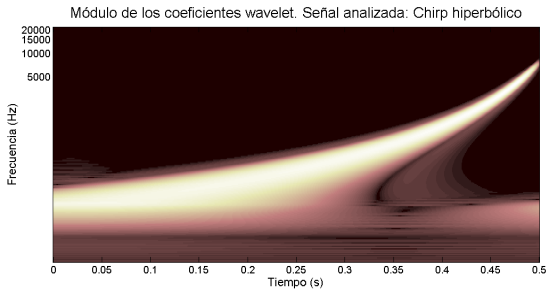
(a) *Espectrograma del tono puro de frecuencia 440Hz. $D = 24$ divisiones por octava.*



(b) *Espectrograma de la señal de 1kHz con excursión sinusoidal de 50Hz. $D = 24$ divisiones por octava.*



(c) *Espectrograma del chirp lineal. $D = 24$ divisiones por octava.*



(d) *Espectrograma del chirp hiperbólico. $D = 24$ divisiones por octava.*

Figura II.16: *Resultados experimentales. Señales no corrompidas (aplicación filtrado). Espectrogramas: (a) Tono A4. (b) FM. (c) Chirp lineal. (d) Chirp hiperbólico.*

tura de la amplitud instantánea desciende a medida que aumenta la relación señal a ruido. Puede parecer que los errores obtenidos en las primeras gráficas de cada figura son elevados (y numéricamente tal vez lo sean, pues llegan a alcanzar, en los peores casos, valores comparables con la propia señal). Sin embargo, la calidad sonora de la señal filtrada resulta ser acústicamente mucho mejor (comparándola con la señal limpia original) de lo que estos resultados parecen reflejar.

Para terminar, en los resultados de recuperación de $f_{ins}(t)$ de las Figuras II.20 a II.22, se ha representado gráficamente el espectro de la señal filtrada. En todos los casos se observa la misma tendencia que en las gráficas de error temporal. A excepción del tono puro de 440Hz, todas las señales presentan un escalograma con un ancho de banda instantáneo bastante marcado, con lo cual incluso eliminando la zona del espectro sobrante, aún queda bastante ruido blanco en la zona espectral de interés como para deslucir el resultado numérico final.

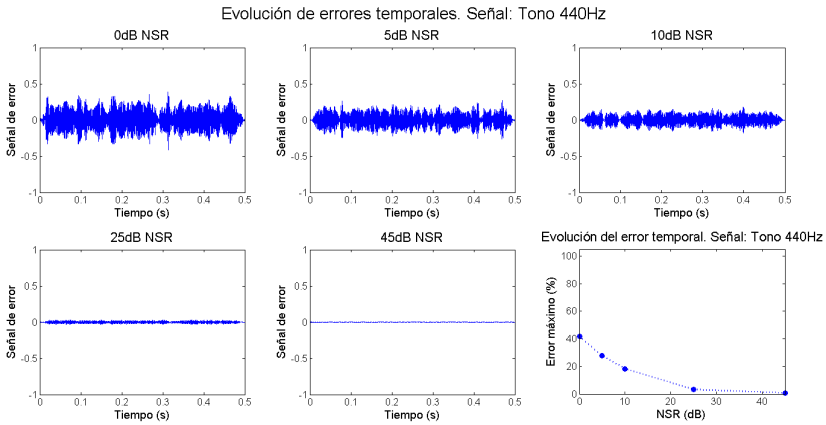


Figura II.17: *Resultados del error temporal para el tono puro de* $440Hz$*. (a)-(e) Resultados para una relación señal a ruido de 0, 5, 10, 25 y 45dB. (f) Gráfica completa del error porcentual cometido.*

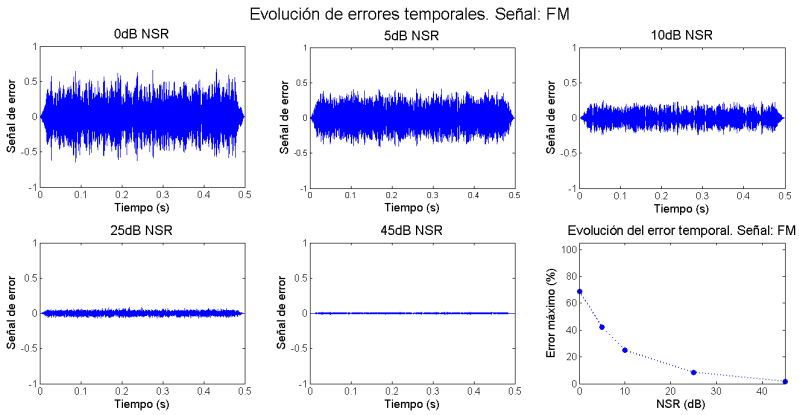


Figura II.18: *Resultados del error temporal para la señal de FM. (a)-(e) Resultados para una relación señal a ruido de 0, 5, 10, 25 y 45dB. (f) Gráfica completa del error porcentual cometido.*

II.d.3. Efectos musicales

La información de módulo y fase coherente tanto en tiempo como en frecuencia que arroja como resultado el algoritmo CWAS es susceptible de ser directamente adaptada o postprocesada de cara a posibles utilizaciones posteriores. Las posibilidades en este campo son múltiples, de modo que a continuación se discutirán cuatro de las eventualmente implementadas, entre las que caben destacar los algoritmos de *pitch shifting* y *time stretching*

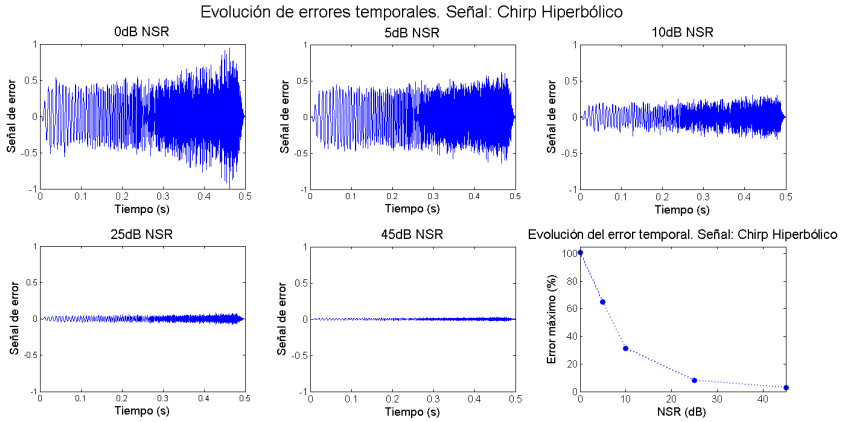


Figura II.19: *Resultados del error temporal para la señal del chirp hiperbólico. (a)-(e) Resultados para una relación señal a ruido de 0, 5, 10, 25 y 45dB. (f) Gráfica completa del error porcentual cometido.*

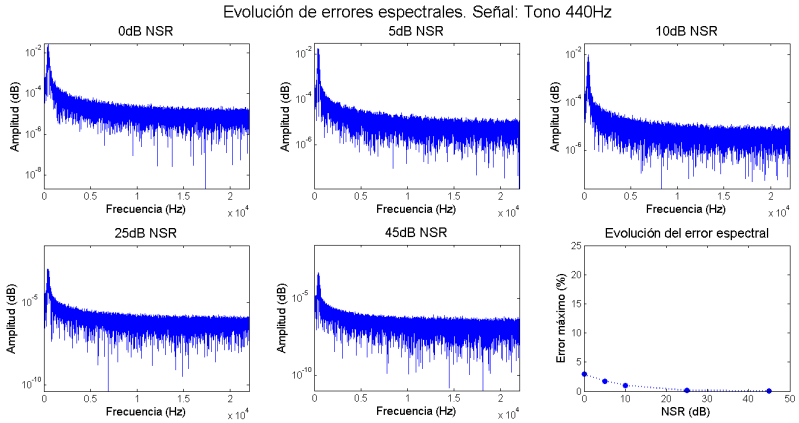


Figura II.20: *Resultados del error espectral para la señal del tono puro de* $440Hz$*. (a)-(e) Resultados para una relación señal a ruido de 0, 5, 10, 25 y 45dB. (f) Gráfica completa del error porcentual cometido.*

a continuación presentados, por su sencillez y la calidad final de los resultados obtenidos. En [70] se han programado hasta 17 efectos digitales diferentes, todos ellos basados en el algoritmo CWAS. La mayoría de estos efectos ha resultado bastante más fácil de programar que mediante los métodos propuestos en el *DAFX* de Udo Zölzer [178], dada la accesibilidad directa a módulos y fases instantáneas de los parciales detectados por nuestro algoritmo. En el soporte informático adjunto a la presente disertación, se incluyen ejemplos varios de los

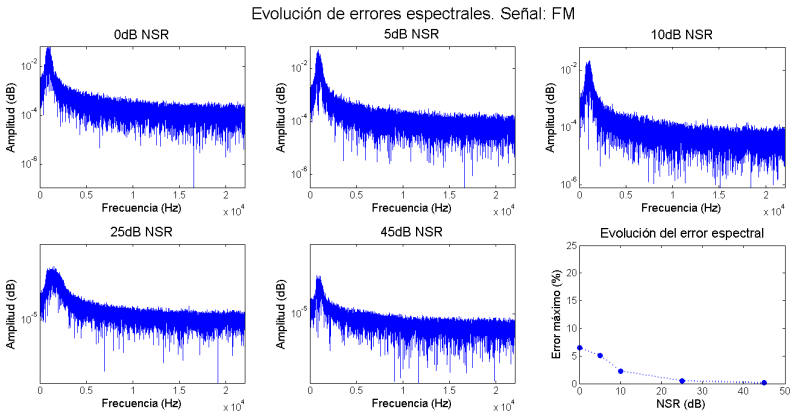


Figura II.21: *Resultados del error espectral para la señal de FM. (a)-(e) Resultados para una relación señal a ruido de 0, 5, 10, 25 y 45dB. (f) Gráfica completa del error porcentual cometido.*

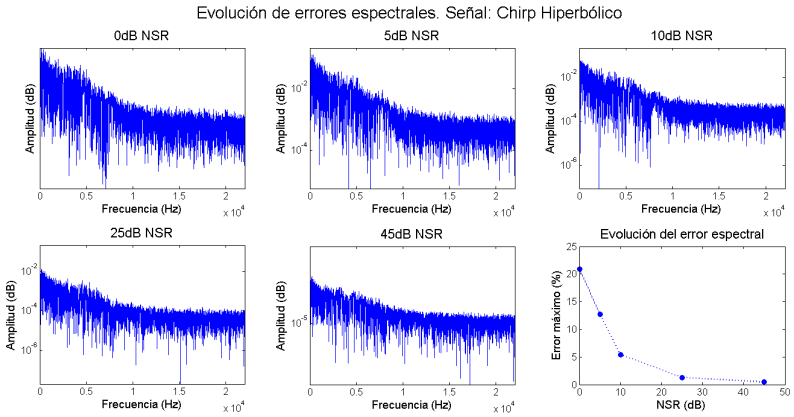


Figura II.22: *Resultados del error espectral para la señal del chirp hiperbólico. (a)-(e) Resultados para una relación señal a ruido de 0, 5, 10, 25 y 45dB. (f) Gráfica completa del error porcentual cometido.*

efectos musicales programados [70].

II.d.3.1. Pitch shifting

Evidentemente, la forma en que disponemos de la información de la señal (es decir, la función compleja que caracteriza a cada parcial de la misma, y por lo tanto su envolvente temporal y su fase instantánea), causa que el efecto del desplazamiento tonal sintético

(afinación) de la señal resulte muy fácil de implementar, al menos en su versión más simple.

El *pitch shifting* consiste en alterar el tono de la señal de audio sin cambiar ni el timbre ni la duración de la misma. Existen muchas formas de llevar a cabo tal efecto, las principales de las cuales se encuentran detalladas en [178]. Pero el algoritmo de pitch shift más simple consiste en multiplicar la fase de la señal por un factor de cambio. Si este factor es mayor que uno, pasaremos a tonos más agudos. Si es menor, los tonos serán más graves.

En nuestro caso, basta con multiplicar la fase instantánea cada parcial de la señal por un coeficiente, sin tocar su envolvente instantánea. Multiplicando cada envolvente por el coseno de su correspondiente fase corregida, tenemos el parcial desplazado. La síntesis aditiva de estos parciales nos permite obtener una señal de gran calidad para desplazamientos en frecuencia relativamente elevados.

Es decir, partiendo de una señal:

$$x(t) = \sum_{i=1}^{N} A_i(t) \cos[\phi_i(t)] \tag{II.16}$$

Su señal desplazada será:

$$x_{psh}(t) = \sum_{i=1}^{N} A_i(t) \cos[\phi_i^{[}r](t)] \tag{II.17}$$

donde:

$$\phi_i^{[r]}(t) = r\phi_i(t) \qquad \forall i \tag{II.18}$$

siendo r la razón de cambio en la fase que provoca el cambio tonal.

En el ejemplo de la Figura II.23, partimos de una señal de flauta ejecutando una nota de frecuencia fundamental $f_0 = 556.6253$Hz, $C\sharp 5$. Su espectro (obtenido mediante la FFT) aparece en azul en la figura. Para cambiar el tono de la señal a una $E5$, $f_1 = 660$Hz, obtenemos un coeficiente de desplazamiento $r = f_1/f_0 = 1.1857$. Aplicando el algoritmo simple propuesto, obtenemos una señal sintética cuyo espectro aparece en rojo en la figura.

Pese a su sencilla implementación, los resultados sonoros de este efecto son de una calidad bastante elevada, incluso para ratios de cambio relativamente grandes. A continuación se muestran una serie de resultados de pitch shifting para los que se ha empleado como base la de la Universidad de Iowa [63] (ver Sección 4.6.3). La idea es realizar un pitch shifting partiendo de una de las señales de las diferentes tablas y terminando en otra. De esta forma es posible comparar los espectros de la señal correspondiente de la base de datos y de la alterada. En las Figuras II.24 y II.25 aparecen los resultados relativos a tres señales tratadas de este modo. En la Figura II.24 se puede ver el resultado del pitch shifting de un Oboe desplazado de $A\#3$ a $B3$. Las líneas verticales negras indican los armónicos co-

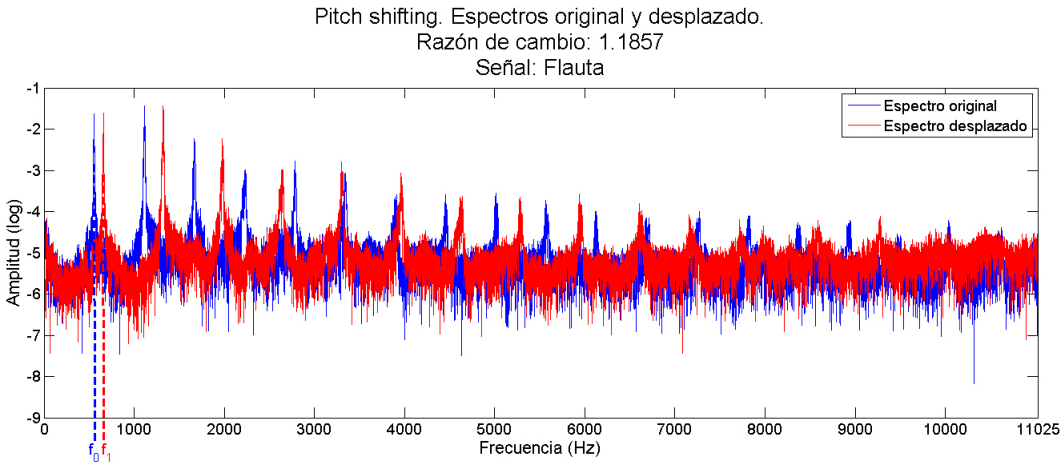


Figura II.23: *Espectros original y desplazado de una señal de flauta de fundamental* $f_0 = 556{,}6253Hz$ *(*$C\sharp 5$*). Fundamental desplazada,* $f_1 = 660Hz$ *(*$E5$*). Factor de desplazamiento (razón de cambio):* $r = f_1/f_0 = 1{,}1857$.

rrectamente situados, mientras que las marcas rojas indican aquellas componentes espurias o desaparecidas del espectro. Dado que el timbre del instrumento viaja con la amplitud y esta no se altera, la señal sintética resulta muy difícil de distinguir de la grabación original. Acústicamente, los resultados son los mismos en las dos señales de la Figura II.25. En esta figura aparecen sendos pitch shifting: de un Clarinete bajo, representado en la Figura II.25(a), con razones de cambio más grandes que en la señal del Oboe, concretamente un salto de $B3$ a $E4$ (5 semitonos) y a $A\#4$ (una escala). Este es el motivo por el que existen un mayor número de componentes deslocalizadas. En el caso de la Figura II.25(b) la situación es incluso peor, ya que se trata de un desplazamiento hacia abajo en frecuencia, con lo cual existe una tendencia intrínseca a perder información de la parte alta del espectro, como se puede apreciar en la figura. Pese a ello, los resultados sonoros siguen siendo aceptables.

Este efecto puede ser mejorado de varias formas. Una de ellas consiste en memorizar una suerte de *envolvente espectral* [35, 70], (para cada instrumento musical, con lo que sería necesario el empleo de una base de datos bastante amplia). Al cambiar de afinación, la envolvente espectral de destino se aplicaría sobre las componentes alteradas, proporcionando un espectro final más parecido al del tono auténtico.

II.d.3.2. Time stretching

Por otro lado, el *time stretching* o cambio en la duración de la señal sin alteración de tono, resulta ligeramente más complicado de programar. Partiendo de los parciales de

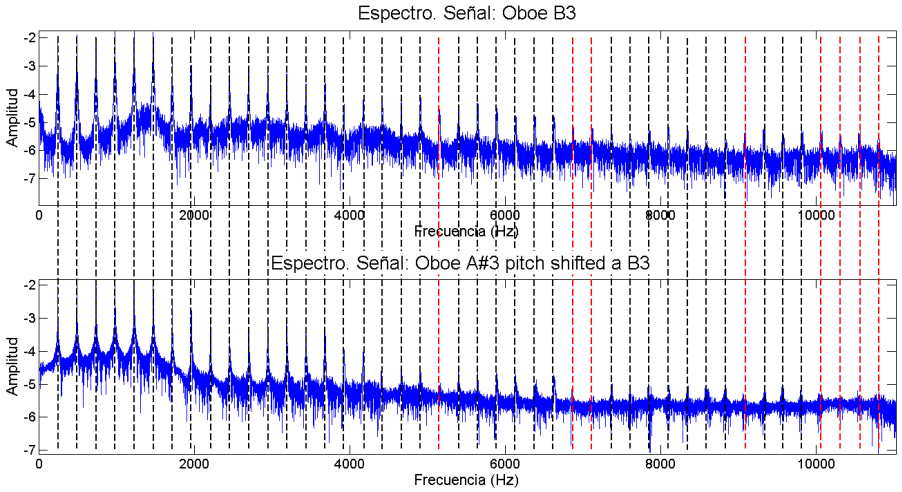


Figura II.24: *Comparativa espectral entre un oboe ejecutando una nota B3 y el pitch shifting de un oboe de A#3 a B3.*

una señal determinada, se pretende extender la duración de la misma mediante el uso de cierto parámetro de control r (concretamente la relación entre las respectivas duraciones de las señales), respetando la afinación de la señal original. Como se detalla en [178], el procedimiento para llevar a cabo el time-stretching de una señal consta de 4 fases principales.

En primer lugar, se buscará el racional p/q más próximo a r, es decir:

$$r \approx \frac{p}{q} \qquad | \quad p, q \in \mathbb{N} \tag{II.19}$$

A continuación, se sobremuestrea cada parcial ρ_n de la señal por p, es decir, se incorporan $p-1$ muestras interpoladas entre dos muestras cualesquiera de cada parcial original. En este caso, se ha utilizado una interpolación cúbica. En Matlab®:

$$\rho_n(t_j) = interp\big\{\rho_n(t_i), p, \text{“cubic”}\big\} \tag{II.20}$$

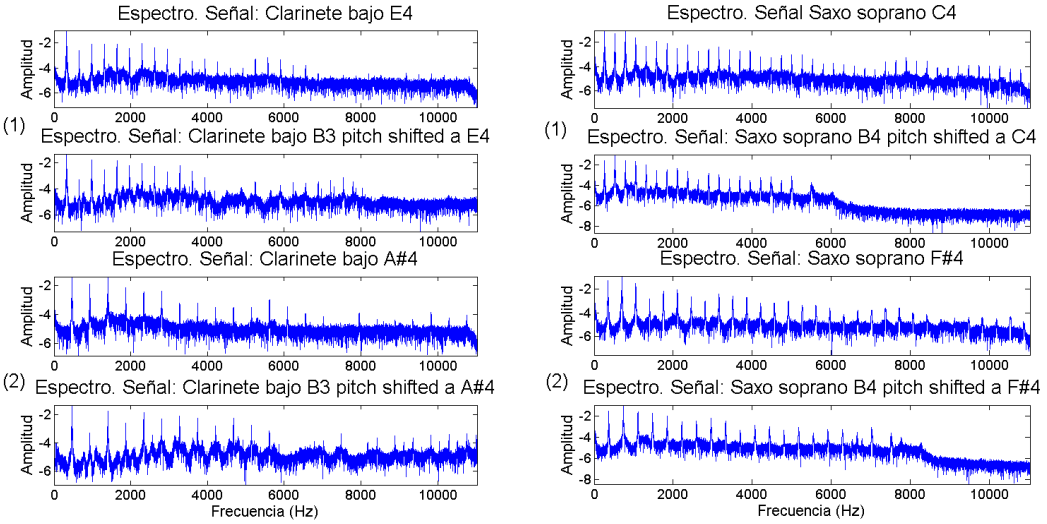
donde el ordinal de t_j es mayor que el de t_i (factor p).

En el tercer paso, se submuestrea cada parcial resultante por un factor q.

$$\rho_n(t_k) = submuestreo\big\{\rho_n(t_j), q\big\} \tag{II.21}$$

donde, evidentemente, el ordinal de t_k es menor que el de t_j (factor q).

Por último, de cada parcial reformado $\rho_n(t_k)$, se extrae la fase resultante $\phi_n(t_k)$ (que



(a) *Pitch shift: clarinete bajo.* (b) *Pitch shift: saxo soprano.*

Figura II.25: *Comparativas espectrales de dos señales con sendos pitch shifting: (a) Señal de Clarinete bajo. Pitch shift up. (1) Señales E4 y B3 a E4. (2) Señales A#4 y B3 a A#4. (b) Señal de Saxo soprano. Pitch shift down. (1) Señales C4 y B4 a C4. (2) Señales F4 y B4 a F4.*

será la fase interpolada del parcial original) y, de forma similar a cómo se procedía en el caso del pitch-shift, se corrige el valor de la misma, en este caso por un factor p/q.

$$\phi'_n(t_k) = \frac{p}{q}\phi_n(t_k) \tag{II.22}$$

Si $A'_n(t_k)$ es la amplitud instantánea del parcial $n - simo$ modificado de la señal, la siguiente síntesis aditiva genera la señal de salida que se busca, $x_{syn}(t_k)$:

$$x_{syn}(t_k) = \sum_{i=1}^{N} \rho_i(t_k) = \sum_{i=1}^{N} A'_i(t_k)\cos[\phi'_i(t_k)] \tag{II.23}$$

En la Figura II.26 aparece representada gráficamente una señal de un clarinete ejecutando una nota $F\#3$, así como los resultados de realizar un time-stretching de la misma mediante ratios $r = 0.69925$, $r = 1.2237$, $r = 1.5733$ y $r = 2$.

Obsérvese al valor de frecuencia fundamental detectado para cada una de estas señales (parte derecha de cada gráfica). Como se deduce de los resultados, la afinación de la señal permanece prácticamente inalterada.

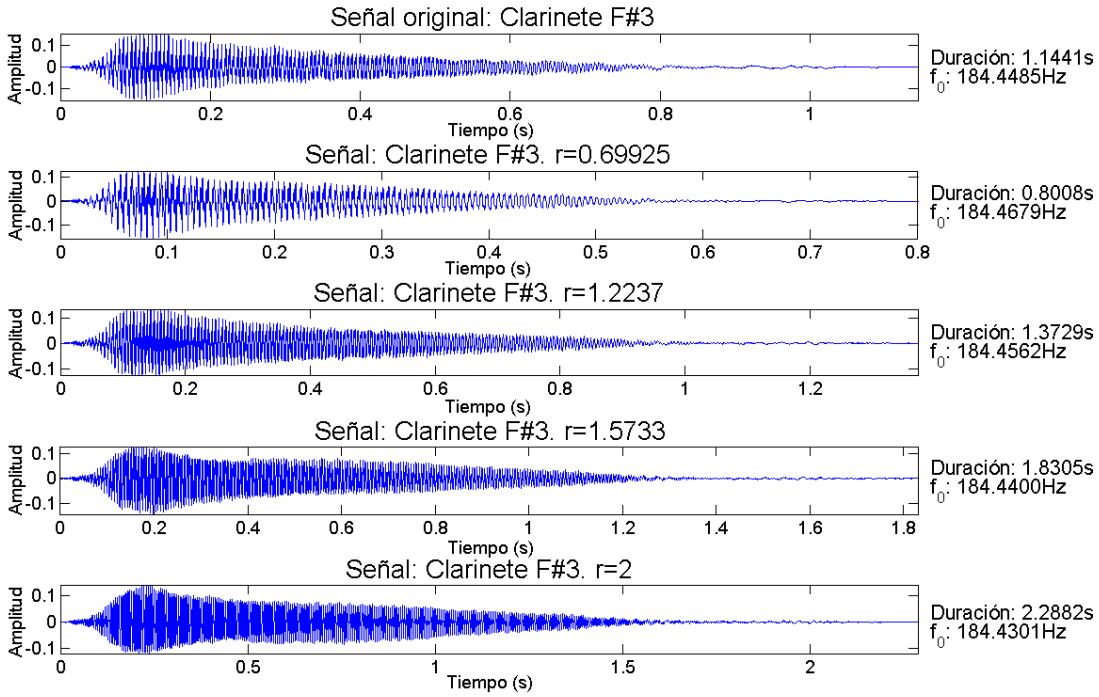


Figura II.26: *Time-Stretching de una señal de Clarinete ejecutando una nota F#3. Arriba: señal original. En orden de representación descendente: r = 0.69925, r = 1.2237, r = 1.5733 y r = 2. A la derecha de cada gráfica aparece la duración final de la señal y la frecuencia fundamental detectada.*

II.d.3.3. Morphing / Síntesis cruzada

Un efecto de sonido especialmente fácil de implementar es el conocido como efecto de *morphing*, entendido como la fusión íntima de dos o más señales diferentes buscando la generación de sonidos nuevos con propiedades híbridas. Este efecto se puede implementar de formas muy diferentes. En una de ellas, se comienza oyendo un sonido y se termina con el otro, resultando la transición un continuo (morphing en el sentido tradicional, equivalente a la inconfundible técnica de imagen en este caso aplicada a señales de audio). En este caso se ha optado por un concepto diferente, una mutación de un sonido dentro de otro [178], que tal vez guarde un mayor paralelismo con la síntesis cruzada (aunque esto depende en gran medida del procedimiento de cambio escogido). El algoritmo queda reflejado en la Figura II.27.

Como se puede ver, se trata de evaluar separadamente cada una de las dos señales utilizando el algoritmo CWAS. Una vez obtenidos fase y módulo de cada parcial en cada una de estas señales, se lleva a cabo la mezcla. Este es un proceso de cuya forma depende

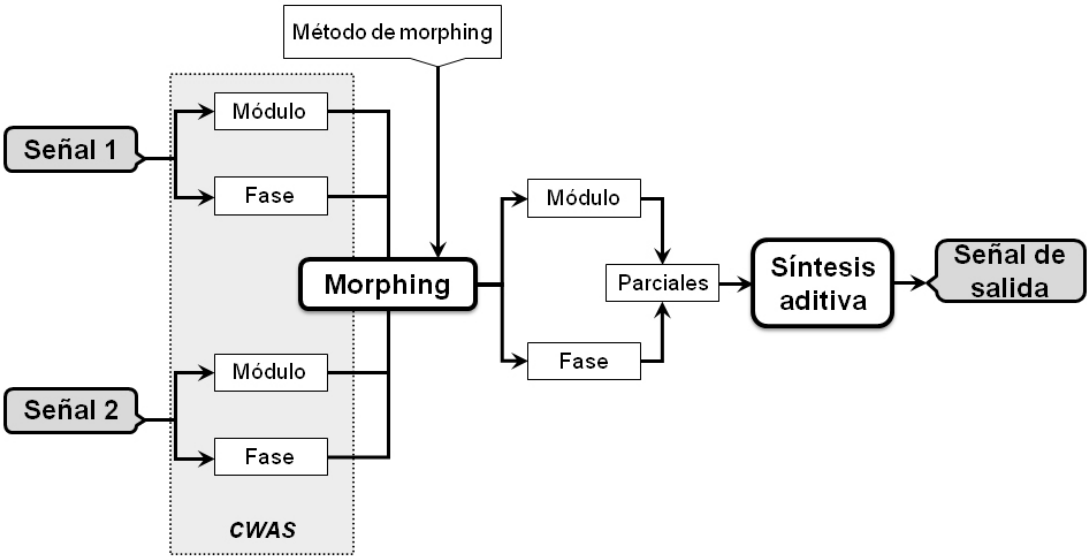


Figura II.27: *Bloque algorítmico general de morphing. Los diferentes resultados se alcanzan en función de la técnica de* morphing *utilizada.*

fuertemente la calidad musical del sonido final, y que puede tener naturalezas muy diferentes. Una posibilidad es utilizar el módulo de los parciales de una señal y la fase de la otra (ya sea globalmente o parcial a parcial). Otra técnica consiste en calcular una amplitud como la suma normalizada de las amplitudes de los parciales fundamental y armónico de cada señal, mientras que la frecuencia instantánea de los mismos se calcula como la semisuma (de modo similar al propuesto en [60]). Sea cual sea el método de morphing seguido, una vez obtenidos módulo y fase para cada parcial, se lleva a cabo el proceso de síntesis aditiva que genera la señal de salida.

En este caso, se lleva a cabo un análisis armónico de ambas señales, tomándose como referencia aquella frecuencial la de la señal que presente un menor número de parciales detectados y como referencia temporal la de la señal más corta. A continuación, comenzando por el parcial fundamental y siguiendo de armónico en armónico se ejecuta el morphing (o la síntesis cruzada) en sí mismo.

Aquí no cabe la evaluación de la calidad del resultado final. El procedimiento pasa por experimentar con diferentes señales de entrada y distintos procesos de mutación, de cara a generar sonidos interesantes. Los resultados finales son sonoramente similares a los que se obtienen por herramientas tales como el programa de síntesis y tratamiento de sonidos `Csound`.

II.d.3.4. Robotización

El efecto clásico de robotización consiste en insertar ceros en los valores de la fase de la FFT antes de la reconstrucción, lo cual fuerza cierta periodicidad en $x(t)$, apareciendo efectos robóticos en lugar de algunas de las variaciones aleatorias propias del sonido original [178] (de forma similar, aunque más compleja, se procede en la conocida técnica PSOLA [119], en busca de características de alto nivel del sonido).

El efecto de robotización basado en CWAS es sutilmente diferente: para eliminar la aleatoriedad propia del sonido original, la fase de cada parcial se reconstruye linealmente a partir de la frecuencia promedio del parcial, es decir, se supone:

$$\phi_{r,n}(t) = 2\pi\overline{f_n}t \tag{II.24}$$

donde $\overline{f_n}$ es la frecuencia característica (frecuencia instantánea promedio) del parcial $n-simo$.

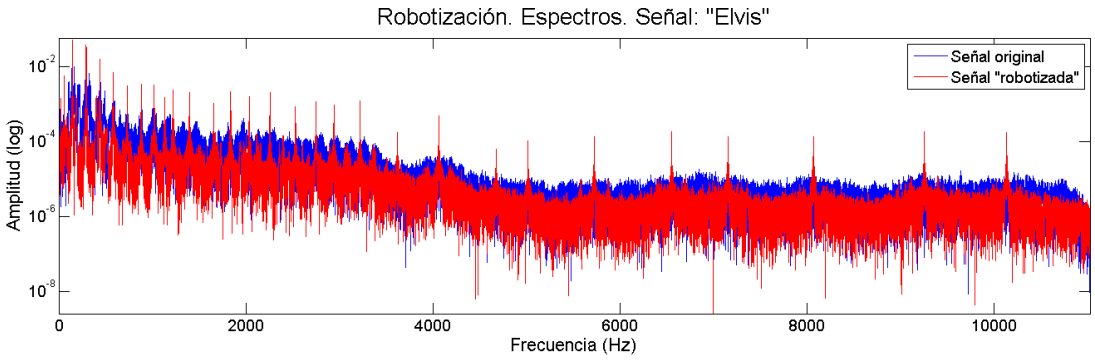


Figura II.28: *Espectros de una señal (en azul) y su correspondiente señal robotizada (en rojo). El espectro de la señal alterada es sensiblemente más regular (periódico) que el de la original, si bien ambos están correlados.*

Utilizando los módulos de los parciales originales y las fases calculadas a través de la Ecuación (II.24), se reconstruye la señal mediante síntesis aditiva:

$$x_r(t) = \sum_{i=1}^{N} A_i(t_k)\cos[m\phi_{r,i}(t)] \tag{II.25}$$

Donde el parámetro m controla el pitch insertado en la señal y por lo tanto la cualidad final del efecto.

En la Figura II.28 se representan el espectro de la señal *"Elvis"* (en azul) analizada en la Sección 5.5.2.1 y su correspondiente señal robotizada (en rojo). Como se puede observar,

la señal alterada presenta un espectro que, sin dejar de ser congruente con el original, ha perdido buena parte de la aleatoriedad entre picos.

Este efecto podría mejorarse, por ejemplo, calculando la frecuencia instantánea promedio de cada parcial en cada ventana del análisis en lugar de emplear el valor promedio global del parcial. De este modo se podría conseguir que el sonido final conserve mejor la entonación del original.

Anexo III

Otros métodos y resultados de separación

"El camino del progreso
no es ni rápido ni fácil".
Marie Curie (1867-1934).
Química y física polaca,
nacionalizada francesa.

En este Apéndice se completan los resultados experimentales relativos a la separación ciega de fuentes (Capítulo 4). Se detallan dos métodos alternativos al presentado en el Capítulo 4: separación por onsets [21] y por distancia armónica. Además se incluye una comparativa de resultados numéricos entre las tres técnicas de separación desarrolladas, y una breve reseña matemática sobre la importancia de la fase en el modelo de la señal, estudio que derivó en el algoritmo de separación de parciales superpuestos propuesto en la Sección 4.7.

III.a. Características y nomenclatura de señales

Como ha visto a lo largo del Capítulo 4, se ha analizado un conjunto bastante amplio de señales mezcladas de dos y tres fuentes. Todas ellas son grabaciones de instrumentos musicales reales, la mayor parte de las cuales se han obtenido de la base de datos de instrumentos musicales de la Universidad de Iowa [63]. Las señales están subsampleadas a $f_s = 22050Hz$

por motivos de ahorro computacional. En los dos primeros algoritmos presentados se trabaja con el número de divisiones por octava estándar; en el tercero se aumenta la resolución aproximadamente por cuatro.

La complejidad y variabilidad de las señales mezcladas ha ido creciendo a medida que lo han hecho las posibilidades del algoritmo, con lo cual ha sido necesario incluir algún tipo de notación con el que el propio nombre de la señal indique las características más destacables de las fuentes empleadas. La notación correspondiente a los instrumentos presentes se incluye en la Tabla III.1. En esta tabla cabe destacar se ha diferenciado, cuando ha sido necesario, entre ejecuciones con y sin vibrato de ciertos instrumentos musicales.

Prefijo de la etiqueta	Instrumento musical
AF	Flauta Alta
AS	Saxo Alto
B	Fagot
BC	Clarinete Bajo
BF	Flauta Baja
C	Clarinete (indeterminado) (*)
Cb	Clarinete en si bemol
Ce	Clarinete en mi bemol
Fv	Flauta (con vibrato)
Fnv	Flauta (sin vibrato)
G	Guitarra (*)
H	Trompa
O	Oboe
P	Piano (*)
S	Saxo (indeterminado) (*)
SS	Saxo Soprano
Tv	Trompeta (con vibrato)
Tnv	Trompeta (sin vibrato)
TrB	Trombón bajo
TrT	Trombón tenor
Tu	Tuba
V	Violín
Vi	Viola

Tabla III.1: *Nomenclatura de instrumentos musicales. Los instrumentos marcados con asterisco (*) no provienen de [63].*

En el nombre final de cada señal mezcla se incluyen cada instrumento presente (siguiendo las abreviaturas de la Tabla III.1) y la nota musical que éste ejecuta. Por ejemplo en la señal $FnvC\#5 + GB4$, la primera fuente sería una flauta tocando una nota $C\#5$ sin vibrato, y

la segunda una guitarra interpretando una $B4$.

III.b. Primera aproximación: onsets

Como se ha adelantado en el Capítulo 4, una primera aproximación para resolver el problema de la separación consiste en emplear únicamente los tiempos de onset que marcan la entrada de las diferentes fuentes presentes en una mezcla [21] como marcadores. Esto dirige la capacidad de separar del algoritmo a aquellos sonidos que no sean simultáneos (esto es, que no sigan un patrón rítmico común), por lo que la técnica puede aplicarse a sonidos interferentes no relacionados rítmicamente. Una utilización práctica de este método podría ser el filtrado selectivo de señales (por ejemplo en audífonos, si bien en este momento se está muy lejos de hacer portable la técnica presentada en esta disertación). El bloque algorítmico del separador basado en onsets aparece representado en la Figura III.1.

Las señales analizadas provienen de las mezclas de diferentes instrumentos musicales ejecutando cada uno de ellos una nota concreta. Más concretamente, se trata de un violín tocando una $G4$, un clarinete $F\#3$, una flauta $C\#5$ (sin vibrato), un saxo $C4$ (con vibrato) y una nota $B4$ de guitarra. En este caso, ninguna de las grabaciones proviene de [63].

En la Figura III.2(a) a III.2(b) se pueden observar los espectrogramas wavelet de dos de las señales analizadas (en concreto, aparecen respectivamente las mezclas de guitarra $B4$ con saxo $C4$ y guitarra $B4$ con clarinete $F\#3$). Aunque la técnica propuesta es en principio independiente del número de fuentes, todas las señales estudiadas son mezclas de dos fuentes. El motivo es que a medida que crece el número de fuentes mezcladas, se hace más importante cierto conocimiento *a priori* de cuantas fuentes hay presentes, de cara a optimizar los resultados del algoritmo. Esto es debido, como se verá más adelante, a que los tiempos de onset son un clasificador bastante limitado del número de fuentes. Este hecho ha desembocado en la posterior degradación del rol jugado por los tiempos de onset y offset dentro de las posteriores versiones del algoritmo.

En los casos presentados en la Figura III.2, resulta visualmente intuitivo distinguir la mayoría de los parciales que pertenecen a cada una de las dos fuentes mezcladas, bien sea por su diferente duración o porque su evolución frecuencial es disímil. Se trata de conseguir el mismo resultado en un proceso automático.

En este primer método se analizan las señales mezcla en un sólo paso (es decir, no hay lectura frame-to-frame ni tracking de parciales). A partir del escalograma se obtienen y cortan los diferentes parciales presentes en la señal (por mínimos, véase la Sección 3.8.1), que serán la base de datos de partida. Existen dos posibilidades para cada uno de ellos: o bien cierto parcial es miembro de una fuente determinada, o se trata de un parcial compartido por más de una fuente.

En este primer acercamiento al problema no se va a abordar la separación de los parciales

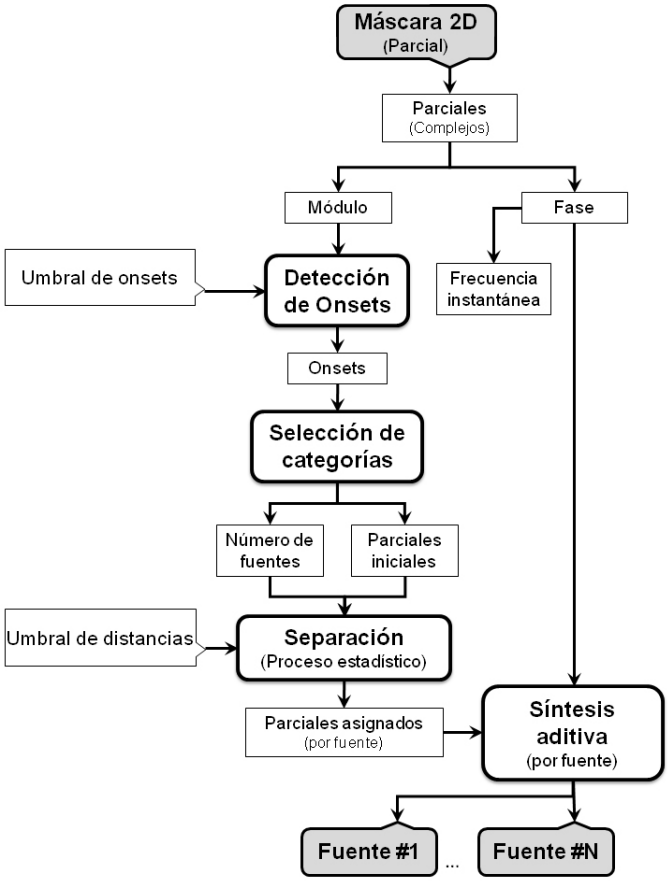
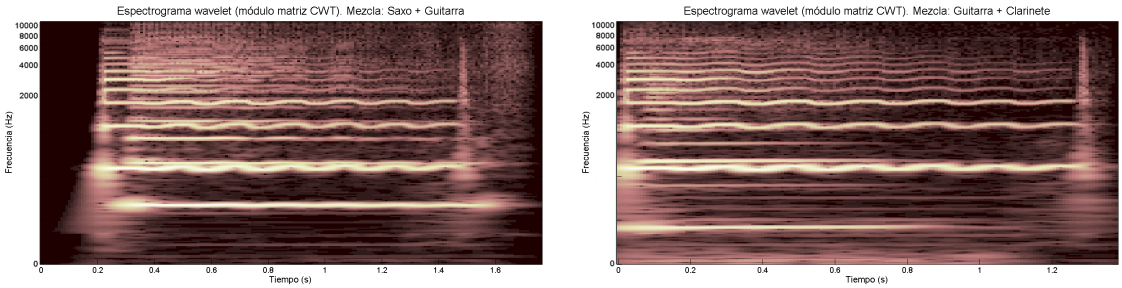


Figura III.1: *Diagrama de bloques del algoritmo de separación monaural basado en análisis estadístico de tiempos de onset.*

superpuestos (Sección 4.7): por el contrario se asignarán a la fuente de mayor presencia o aquella con la que guarden un mayor parecido. Esto, como se demostrará, provoca la aparición de un error por interferencia más evidente en una de las fuentes y es una de las principales limitaciones de la técnica.

Con el fin de separar las fuentes, es necesario establecer algún *criterio de semejanza* entre los parciales, el cual se puede extraer de la frecuencia y la amplitud instantáneas de las diferentes componentes detectadas. El objetivo es obtener algún tipo de *patrón de similitud* entre los parciales de la misma fuente. Este patrón se obtiene mediante la evaluación de las diferencias en la evolución de envolvente y frecuencia instantánea de cada parcial con respecto a todos los demás [168].

Los distintos parciales detectados en una señal presentan una gran variedad de valores



(a) *Espectrograma. Señal: mezcla de guitarra y saxo.*

(b) *Espectrograma. Señal: mezcla de guitarra y clarinete.*

Figura III.2: *Módulo de los coeficientes wavelet (espectrograma wavelet) de las cuatro señales analizadas. (a) Señal mezcla de guitarra y saxo. (b) Señal mezcla de guitarra y clarinete.*

de amplitud, lo cual imposibilita la utilización directa de esta información en la búsqueda de similitudes. Sin embargo, escalando cada parcial por su valor medio, las envolventes resultantes sí comienzan a resultar comparables (efecto conocido como Modulación de Amplitud Común, en el que se ha incidido en el Capítulo 4).

Por otro lado, debido a la propia naturaleza del sonido, la frecuencia instantánea *per se* no presenta tantas variaciones en promedio como las amplitudes, y por lo tanto resultará un parámetro de comparación más relevante. Sin embargo, dado que en pasos posteriores se habrá de comparar la evolución por pares de estas frecuencias, también resulta razonable que sean normalizadas previamente (con lo cual las frecuencias de todos los parciales oscilarán en torno a 1, y se podrán calcular diferencias de un modo directo).

En la Figura III.3 se ha representado la evolución de la envolvente y de la frecuencia instantánea normalizadas de los tres parciales más importantes de la señal mezcla flauta+clarinete. Se puede apreciar que la frecuencia instantánea de estos parciales evoluciona de forma muy similar, mientras la evolución en las envolventes no lo es tanto.

La variación de la frecuencia instantánea en puntos de baja amplitud puede resultar un problema (sobre todo en los parciales de mayor frecuencia, donde el número de puntos problemáticos puede resultar importante). Para evitar esta posible fuente de error, en lugar de trabajar directamente con las magnitudes instantáneas, se llevará a cabo un proceso de suavizado que eliminará las variaciones superfluas, manteniendo la información relevante prácticamente intacta. Este filtrado se lleva a cabo de la misma forma con las amplitudes, por motivos similares. En la Figura III.3, se ha representado la información ya filtrada. En cualquier caso, dado que los problemas podrían sobre todo aparecer en parciales (o partes de parciales) de baja amplitud, cabe esperar que el efecto residual de estos errores acumulativos

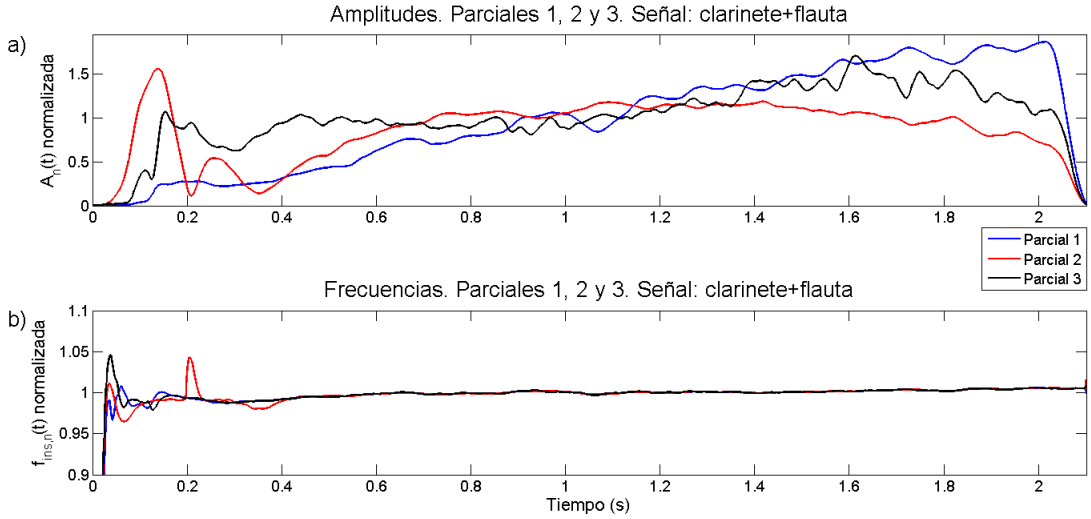


Figura III.3: *Evolución temporal de amplitudes y frecuencias de los tres parciales más energéticos de la señal mezcla de flauta y clarinete.*

sea mínimo.

III.b.1. Distancias modular y frecuencial

El método de comparación de parecido entre los diferentes parciales va a ser la evaluación de una distancia promedio, compuesta a su vez por la combinación lineal de dos distancias diferentes: la distancia modular y la distancia frecuencial. Estas pueden ser definidas a partir del error cuadrático medio [168], como:

$$d_m(i,j) = \frac{1}{\|\Delta t\|} \sum_{t=0}^{\Delta t} \left[\frac{m_i(t)}{\bar{m}_i} - \frac{m_j(t)}{\bar{m}_j} \right]^2 \tag{III.1}$$

$$d_f(i,j) = \frac{1}{\|\Delta t\|} \sum_{t=0}^{\Delta t} \left[\frac{f_i(t)}{\bar{f}_i} - \frac{f_j(t)}{\bar{f}_j} \right]^2 \tag{III.2}$$

donde $\|\Delta t\|$ es la duración temporal de la señal (en muestras), $m_{i,j}(t)$ y $f_{i,j}(t)$ son las amplitudes y frecuencias instantáneas de los parciales $i-simo$ y $j-simo$, mientras que $\bar{m}_{i,j}$ y $\bar{f}_{i,j}$, son los módulos y frecuencias promedio de los mismos.

La distancia global entre parciales es una suma ponderada de estas dos distancias, es decir:

$$d(i,j) = w_m d_m(i,j) + w_f d_f(i,j) \tag{III.3}$$

donde w_m y w_f son, respectivamente, los pesos asociados a los errores cuadráticos promedio modular y frecuencial. Los valores concretos que se han utilizado para el análisis de las señales mostrado más adelante son $w_m = 0.1$ y $w_f = 0.9$, si bien los resultados cualitativos de separación no dependen significativamente del valor exacto de estos parámetros.

La Ecuación (III.3) es una distancia entre pares. Si la señal presenta un total de P parciales, aplicando la Ecuación (III.3) se obtiene una matriz de distancias D de tamaño $P \times P$ con todas las distancias cruzadas entre parciales por encima y por debajo de la diagonal principal (la matriz, obviamente, es simétrica) y ceros en la misma:

$$D = \begin{pmatrix} 0 & d(1,2) & \dots & d(1,P) \\ d(2,1) & 0 & \dots & d(2,P) \\ \vdots & \vdots & \dots & \vdots \\ d(P,1) & d(P,2) & \dots & 0 \end{pmatrix}$$

La mayor dificultad consiste en cómo interpretar esta información adecuadamente de cara a separar las fuentes, asignando correctamente los parciales que correspondan a cada una de ellas. El problema es que, *a priori*, ni siquiera se conoce el número de fuentes presente en la mezcla. En este caso tampoco se sabe si un parcial concreto cualquiera se corresponde con alguna fuente determinada o bien es un parcial compartido. Estas condiciones de contorno implican que sólo se puede proceder intentando aplicar la estadística y la lógica, lo cual supone a su vez un sesgo natural a la dependencia de la señal en los resultados finales.

III.b.2. Tratamiento estadístico. Onsets

En este punto se dispone de un indicador de similitud entre parciales, las distancias $d(i,j)$. Es evidente que los parciales pertenecientes a una misma fuente se distinguen por presentar una distancia total que tiende a cero. De este modo se hace necesario agrupar los parciales en tantas familias o categorías *disjuntas* como fuentes haya en la señal, haciendo que los miembros respectivos de cada clase presenten un error mínimo en su distancia evaluada por pares. Partiendo de la expresión [168]:

$$min\left[\frac{1}{\|\mathbf{S}_1\|}\sum_{i,j\in\mathbf{S}_1} d(i,j) + \frac{1}{\|\mathbf{S}_2\|}\sum_{k,l\in\mathbf{S}_2} d(k,l) + \dots\right] \tag{III.4}$$

La Ecuación (III.4) supone que las diferentes fuentes presentes son disjuntas (es decir, que no tienen espectros superpuestos), y en ella $\mathbf{S}_1$, $\mathbf{S}_2$,... etc., son los diferentes conjuntos de parciales correspondientes con cada fuente, $\|S_{1,2}\|$ es la cardinalidad de cada conjunto, $\mathbf{S} = \mathbf{S}_1 \bigcup \mathbf{S}_2$ es el conjunto completo de parciales detectados y $\mathbf{S}_1 \bigcap \mathbf{S}_2 = \emptyset$.

En este punto, la solución ideal pasaría por calcular todas las posibles permutaciones de

la Ecuación (III.4), y escoger la mejor. Sin embargo, el número de posibilidades resulta muy elevado, de modo que será necesario encontrar algún atajo que permita alcanzar la solución correcta en un tiempo más razonable. Por ello, intentando además mantener al máximo el conjunto de señales para el que se pueda obtener un resultado de separación satisfactorio, no se trabajará con el conjunto total de parciales detectados sino con el subconjunto de los más energéticos.

Por otro lado, dado que una de las incógnitas más relevantes del problema consiste en determinar el número de fuentes presentes en la mezcla, se ha utilizado la siguiente hipótesis:

- *El conjunto de parciales relacionado con la misma fuente presenta* aproximadamente *el mismo tiempo de onset.*

Para cada parcial, antes del correspondiente onset y después del offset, la amplitud resulta despreciable (es decir, desciende por debajo de cierto umbral). Atendiendo de nuevo a la Figura III.3, los tres parciales representados forman parte de la misma fuente (en este caso, la flauta). Como se puede apreciar, si bien las envolventes no son tan similares como para poder resultar un parámetro adecuado de decisión, los tiempos de onset sí son bastante parecidos entre sí.

Se ha utilizado un algoritmo de localización y clasificación de onsets basado en el presentado en la Sección 4.5.2 [23] pero más simple (puesto que no es necesaria una precisión excesiva en la localización del onset). Este algoritmo calcula el tiempo de entrada y salida de cada parcial detectado, y los agrupa en categorías por medio de un parámetro de umbral θ_o (véase de nuevo la Figura III.1). Es evidente que esto limita la utilidad del algoritmo pues se hace necesaria una separación crítica en los eventos de entrada de las fuentes.

El algoritmo de onsets trata de agrupar parciales que se correspondan claramente a fuentes diferentes, asumiendo que un mismo instrumento produce una única nota aislada en cada instante de tiempo. Sin embargo, los tiempos de onset y offset son de algún modo parámetros dependientes de la señal, como se demostrará gráficamente más adelante, de modo que pueden ser utilizados únicamente para encontrar los conjuntos de candidatos firmes a miembros de cada fuente presente en la mezcla. A mayor energía del parcial, mejor definirá éste las características de envolvente y frecuencia de su fuente asociada (de ahí la limitación energética mencionada anteriormente). Una vez detectados los conjuntos iniciales de parciales energéticos más claramente pertenecientes a fuentes distintas, se procede a evaluar la distancia entre estos parciales y todos los demás detectados y no asignados utilizando las Ecuaciones (III.1), (III.2) y (III.3). Si la distancia está por debajo de cierto umbral θ_d (consúltese de nuevo la Figura III.1), se considera que el parcial pertenece a la fuente en cuestión. Así, un mismo parcial puede pertenecer a varias fuentes (síntoma inequívoco de espectros superpuestos). En este caso, el parcial será asignado a la señal más presente o en su defecto a aquella que se encuentre situada a menor distancia.

Tomando las frecuencias (escalas) relacionadas con cada uno de los parciales asignados a la fuente $k-sima$, se puede construir una máscara bidimensional $M(a_{i,k},t)$, la cual puede a su vez ser utilizada para sintetizar la señal separada $\bar{s}_k(t)$ (compuesta por P_k parciales), partiendo de los coeficientes wavelet de la mezcla:

$$\bar{s}_k(t) = \sum_{i=1}^{P_k} \rho_i(t) = \sum_{i=1}^{P_k} M(a_{i,l},t) \circ W_x(a,t) \tag{III.5}$$

donde el operador $\circ$ es el *producto de Hadamard* o producto elemento por elemento (equivalente al operador $.*$ de Matlab®), mientras que $W_x(a,t)$ son los coeficientes wavelet complejos de la señal original (mezcla) y $a_{i,k}$ las escalas relacionadas con el parcial $i-simo$ de la fuente $k-sima$.

III.b.3. Resultados, limitaciones y valoración

Este algoritmo ha sido puesto a prueba con un pequeño conjunto de señales mezcladas de forma sintética, correspondientes a los diferentes instrumentos musicales ejecutando notas distintas mencionados anteriormente. Los resultados numéricos de calidad de la separación aparecen en las Tablas III.2 a III.4. Los nombres de las señales analizadas están codificados siguiendo la nomenclatura de la Tabla III.1.

Señal	SDR_1	SDR_2	$\overline{SDR}$
CF#3+FnvC#5	2.5567	15.4105	8.9836
CF#3+GB4	16.6164	23.7654	20.1909
GB4+SvC4	9.7176	9.3438	9.5387

Tabla III.2: *Separación por onsets. Resultados numéricos: Parámetro SDR (dB).*

Señal	SIR_1	SIR_2	$\overline{SIR}$
CF#3+FnvC#5	39.9546	19.8616	29.9081
CF#3+GB4	68.2717	43.1652	55.7184
GB4+SvC4	49.2039	19.3753	34.2896

Tabla III.3: *Separación por onsets. Resultados numéricos: Parámetro SIR (dB).*

Señal	SAR_1	SAR_2	$\overline{SAR}$
CF#3+FnvC#5	2.5579	17.3854	9.9717
CF#3+GB4	16.6164	23.8158	20.2161
GB4+SvC4	9.7182	9.8478	9.7830

Tabla III.4: *Separación por onsets. Resultados numéricos: Parámetro SAR (dB).*

Con estos resultados, los valores promediados de los parámetros para las señales analizadas son:

- $\overline{SDR_o}$= 12.9044 dB.
- $\overline{SIR_o}$= 39.9720 dB.
- $\overline{SAR_o}$= 13.3236 dB.

Como figuras de mérito, se muestran las formas de onda correspondientes a las señales cuyo espectrograma se ha presentado en la Figura III.2.

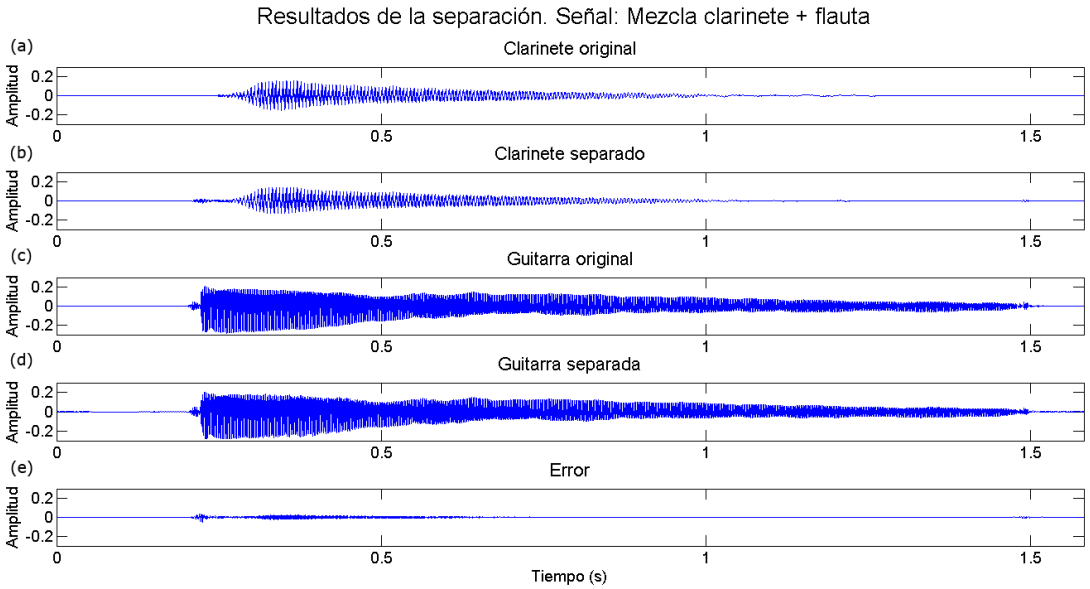


Figura III.4: *Formas de onda correspondientes a los resultados de la separación de la señal mezcla de guitarra y clarinete. (a) Clarinete original. (b) Guitarra original. (c) Clarinete separado. (d) Guitarra separada. (e) Error.*

Mientras la separación temporal entre las entradas sea suficiente (algo relativamente fácil de asumir) y menor número de parciales mezclados se encuentren, los resultados ofrecidos por esta técnica son bastante satisfactorios. Como muestra de este hecho, en la Figura III.4 se presentan los resultados de la separación de clarinete y guitarra. Su espectrograma es el de la Figura III.2(b). Como se puede ver por las formas de onda, excepto algo de información de las bandas de alta frecuencia (donde la superposición de parciales es prácticamente inevitable) el resto de parciales ha sido adecuadamente asignado a las fuentes correctas.

Sin embargo, cuando la zona del espectro superpuesto es más importante respecto del total, o cuando algún parcial suficientemente energético está montado sobre un parcial de otra fuente o es asignado erróneamente por la umbralización de los tiempos de onset, la calidad final del resultado se resiente. En la Figura III.5 se ha representado el resultado de la separación de la señal de flauta y clarinete. Como se puede observar, la gran superposición de parciales de alta frecuencia ocasiona que gran parte de la señal del clarinete se asigne a la señal de mayor presencia, la flauta, ocasionando un error perfectamente mensurable y audible. De hecho, la mitad de la energía del clarinete se ha asignado por este motivo a la flauta, si bien el resultado sonoro resulta pese a todo aceptable. En la Sección III.b.4 se presentan resultados de separación adicionales empleando este método.

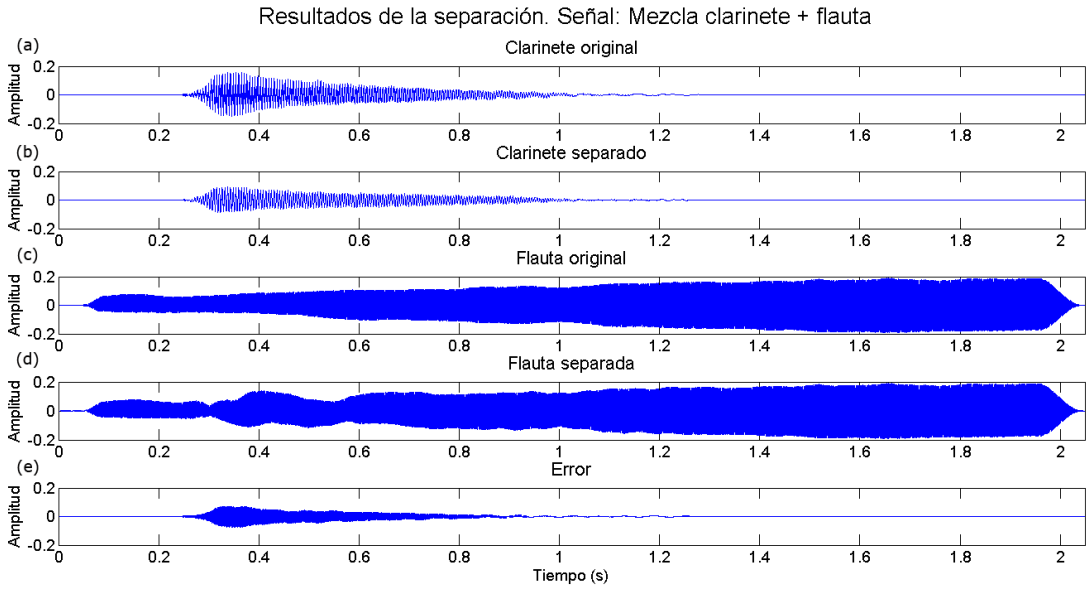


Figura III.5: *Formas de onda correspondientes a los resultados de la separación de la señal mezcla de flauta y clarinete. (a) Clarinete original. (b) Flauta original. (c) Clarinete separado. (d) Flauta separada. (e) Error.*

Las limitaciones de esta técnica son evidentes. En primer lugar, las fuentes que evolucionen de forma síncrona no serán adecuadamente separadas. Por añadidura, el concepto de sincronía está relajado en este caso debido al uso del umbral de onsets θ_o. Como ilustración de este hecho, en la Figura III.6 se ha representado el escalograma wavelet de la señal de saxo+flauta.

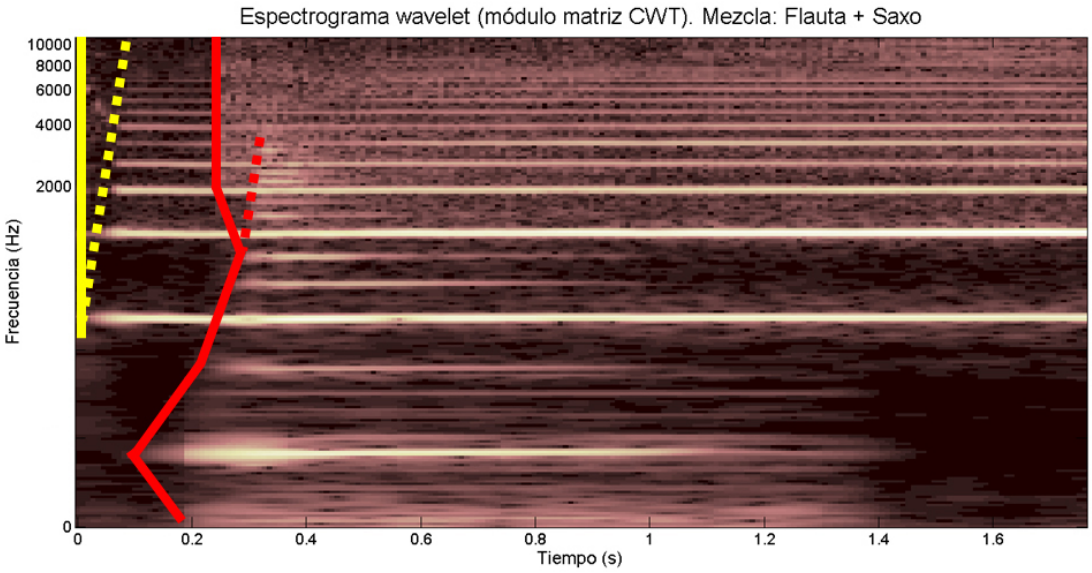


Figura III.6: *En amarillo y rojo, trazos aproximados de los tiempos de onset para parciales de flauta y saxo, respectivamente, marcados sobre su espectrograma wavelet.*

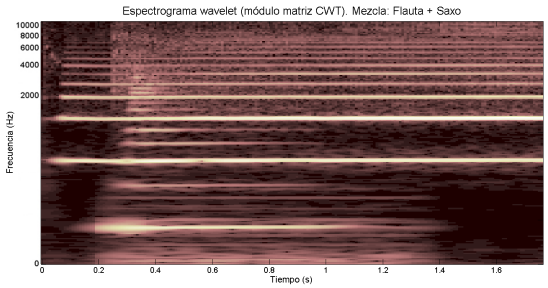
En la figura, en trazo amarillo y rojo se han remarcado los onsets de cada uno de los instrumentos. Como se puede observar, el tiempo de onset de la flauta (amarillo) es razonablemente estable (si bien esta estabilidad es cuestionable y sólo se alcanza mediante la adecuada umbralización. La trayectoria discontinua parece más aproximada al comportamiento real de la señal). En rojo, los onsets de los parciales del saxo. La variabilidad resulta en este caso mucho mayor. Englobar estos parciales como miembros de una misma fuente y no añadir por error alguno correspondiente a la señal de la flauta depende en gran medida de la flexibilidad en la definición de *simultáneo* que se tome.

El hecho de estudiar la señal de entrada en un único paso (es decir, no frame-to-frame) supone una nueva fuente de error: los parciales serán estructuras con un ancho de banda determinado, pero su duración temporal coincidirá *siempre* con la duración total de la señal, independientemente de la duración de la fuente involucrada. Esto genera la aparición de espurios finales (de baja energía pero igualmente audibles) cuando la amplitud de la fuente separada decae por debajo de cierto umbral.

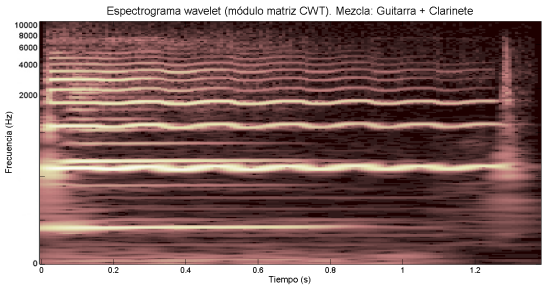
III.b.4. Resultados gráficos adicionales

En la Sección anterior se han presentado los resultados numéricos y gráficos obtenidos en la separación de fuentes monaurales por tratamiento estadístico de onsets. El conjunto total de señales analizadas fue de seis, si bien en este momento sólo se dispone de las señales separadas de tres de ellas. Estas tres señales son de las que se han obtenido los parámetros de calidad estándar presentados en las tablas III.2 a III.4. Sin embargo, en [21] se presentan los resultados gráficos correspondientes a cuatro de las seis señales analizadas, dos de los cuales han sido incluidos en la citada Sección III.b.3.

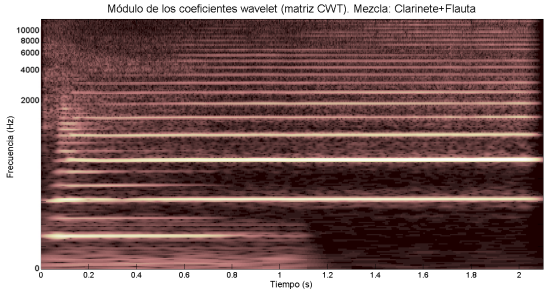
En la Figura III.7 se han representado los espectrogramas wavelet de las cuatro señales que completan el conjunto de seis junto a las dos presentadas en la Figura III.2. Como se puede apreciar, el nivel de superposición de parciales es bastante elevado en la zona de frecuencias medias/altas, en todos los casos. Más aún, las señales mezcla de flauta más saxo, Figura III.7(a) y flauta más clarinete, Figura III.7(c), presentan superposición en parciales con gran presencia energética. El algoritmo no incluye la separación de estos parciales.



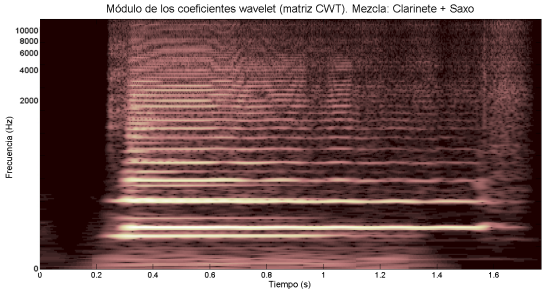
(a) *Espectrograma. Señal: mezcla de flauta y saxo.*



(b) *Espectrograma. Señal: mezcla de guitarra y clarinete.*



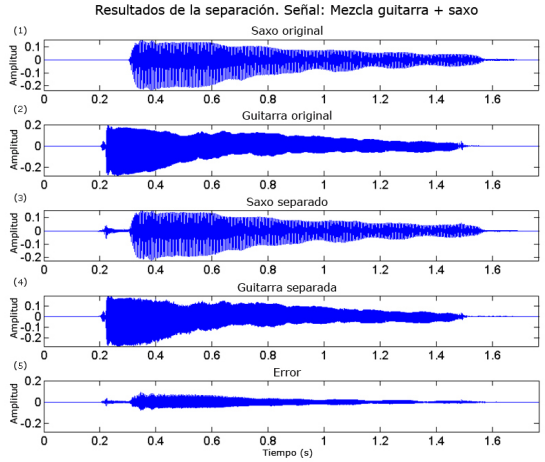
(c) *Espectrograma. Señal: mezcla de clarinete y flauta.*



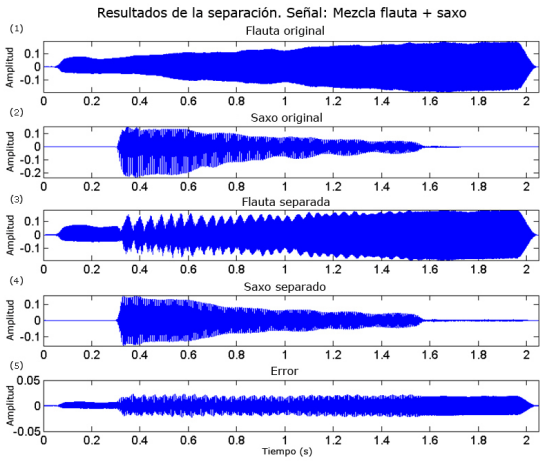
(d) *Espectrograma. Señal: mezcla de clarinete y saxo.*

Figura III.7: *Módulo de los coeficientes wavelet (espectrograma wavelet) de las restantes cuatro señales analizadas. (a) Señal mezcla de flauta y saxo. (b) Señal mezcla de guitarra y clarinete. (c) Señal mezcla de clarinete y flauta. (d) Señal mezcla de clarinete y saxo.*

Para terminar, en las Figuras III.8(a) y III.8(b) se incluyen los resultados gráficos de separación correspondientes respectivamente a las señales mezcla de guitarra más saxo y flauta más saxo (formas de onda). En cada subfigura, los ejes no son iguales, de modo que la correspondencia visual de las señales no es directa.



(a) *Mezcla de guitarra y saxo.*



(b) *Mezcla de flauta y saxo.*

Figura III.8: *Resultados de separación por el método de onsets. (a) Formas de onda correspondientes a los resultados de la separación de la señal mezcla de guitarra y saxo. (1) Saxo original. (2) Guitarra original. (3) Saxo separado. (4) Guitarra separada. (5) Error. (b) Formas de onda correspondientes a los resultados de la separación de la señal mezcla de flauta y saxo. (1) Flauta original. (2) Saxo original. (3) Flauta separada. (4) Saxo separado. (5) Error.*

Cuando algún parcial de contenido energético relativamente grande no es correctamente separado, la correspondiente señal de error tiende a ser mayor, Figura III.8(b)-5. Esto lleva a la conclusión de que, de cara a alcanzar resultados de separación de calidad indiscutible, se hace necesario separar estos parciales solapados.

III.c. Separación por armonía y distancia armónica

Como se ha explicado, la técnica inicial de separación tropieza con tres problemas evidentes: un tratamiento temporal de la información excesivamente simplista, la poca fiabilidad del tiempo de onset como único parámetro de control y por último la separación de los parciales superpuestos (con mucho el desafío más complicado de resolver). En esta segunda aproximación se trata de superar los dos primeros problemas. En cuanto al tercero, se ha buscado eludirlo de un modo directo pero no selectivo. Es decir, la mayor parte de los proble-

mas de parciales superpuestos se da a frecuencias relativamente elevadas y con parciales de energía relativamente baja. Si estos parciales no son tenidos en cuenta en la reconstrucción, la energía erróneamente asignada a una fuente dada disminuirá. Esto tiende a mejorar los resultados *numéricos* de la separación, si bien hace que las señales obtenidas pierdan buena parte del *color* característico de los originales aislados. Sin embargo, en esta aproximación no se ha empleado un algoritmo de búsqueda de fundamentales tan completo como el que se ha expuesto en el Capítulo 4, como se explicará más adelante.

El segundo algoritmo de separación completo aparece representado esquemáticamente en la Figura III.9. Cabe destacar que en este caso el análisis de la información tiempo–frecuencia se lleva a cabo en un proceso frame-to-frame anidado (véase Sección 3.9.3: tamaño de la trama, 4095 muestras; tamaño del segmento, 256 muestras). Por otro lado, se ha sustituido el elemento indicador de fuentes, reemplazándose los tiempos de onset por una *distancia armónica*, mucho más eficiente. El uso de este parámetro implica que los instrumentos musicales presentes en la mezcla deben ser necesariamente armónicos[1], lo que conduce a resultados de menor calidad ante instrumentos inarmónicos dado que los parciales superpuestos se asignarán a una única fuente en función de un *parámetro de distancia*, como se verá.

Como se puede apreciar en la Figura III.9, se ha umbralizado el nivel de energía mínimo para que los parciales del escalograma sean tenidos en cuenta en el proceso de tracking:

$$\theta_e = \theta_{ea} \cdot \sum_{i=1}^{L} max\left[\|W_x(k,t_i)\|^2\right] \tag{III.6}$$

siendo L la duración total de la señal y θ_{ea} un umbral semi-adaptativo que puede alcanzar dos valores diferentes, θ_{ea} =0.08 y θ_{ea} =0.04, en función de la presencia o no de un onset. El onset se localiza en este caso de forma aproximada dentro de cada segmento de 256 muestras, analizando el valor RMS del escalograma[2] (marcado por un aumento en su nivel medio respecto al impuesto por el ruido de fondo). Este procedimiento se hace necesario para evitar el tracking de parciales energéticamente poco importantes al inicio de la señal, optimizando el tiempo de proceso. En la Ecuación (III.6), el sumatorio representa el escalograma *global* de

[1] *Dado que la música occidental favorece la escala temperada de doce tonos, los intervalos musicales comunes presentan relaciones tonales muy próximas a cocientes de valores enteros [110] (≈ 2/3, 4/3, 5/3, 5/4, etc.) Como consecuencia, un gran número de armónicos de una fuente estarán superpuestos con armónicos de las demás en una mezcla. Por ejemplo, en una relación de quinta justa (3/2), una fuente tiene los armónicos pares superpuestos, y la otra los múltiplos de tres.*

[2] *En efecto, el transitorio asociado al ataque en una señal musical tiende a extender la información a lo largo del eje de frecuencias. Esto hace posible detectar los onsets midiendo la anchura de los parciales en escala, o lo que es equivalente, el nivel RMS promedio del escalograma. Este efecto es tanto más evidente cuanto más abrupto sea el ataque de la señal, pero resulta perfectamente visible en la mayoría de los espectrogramas wavelet presentados a lo largo de esta disertación. Como caso extremo, puede tomarse el de la batería (Figura I.10).*

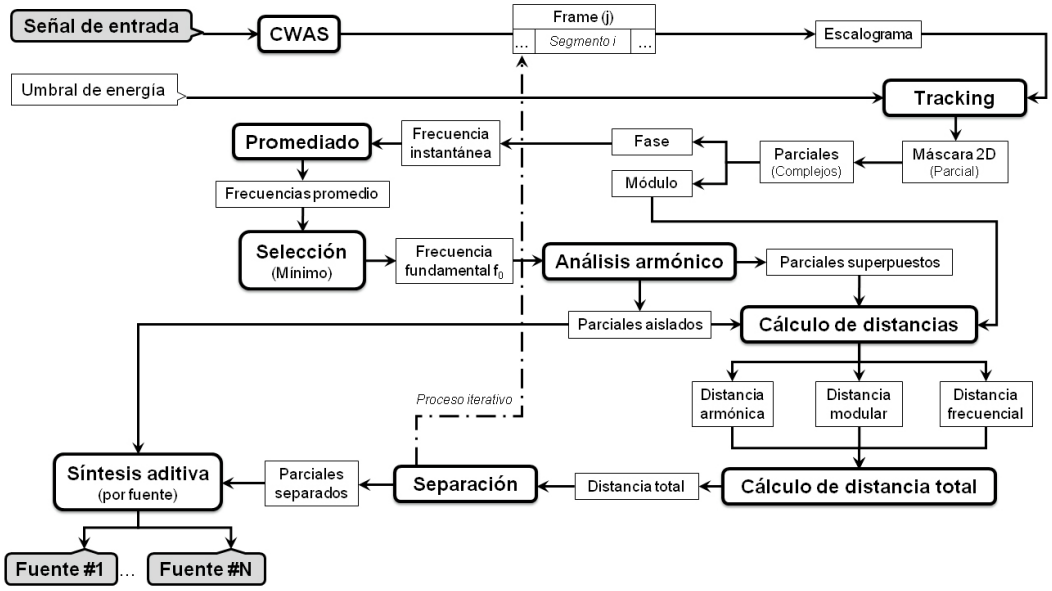


Figura III.9: *Bloque esquemático del algoritmo CWAS aplicado a separación por distancias.*

$x(t)$, lo que supone que la matriz de coeficientes será leída dos veces: la primera para obtener el escalograma y la segunda para realizar el proceso frame-to-frame. Esta falta de eficiencia puede ser suplida utilizando un umbral energético local o adaptativo, pero la mejora se ha descartado finalmente, asumiendo que ésta técnica es un nuevo paso previo al algoritmo final, como se ha visto en la Sección 4.7.

En cada trama de $x(t)$ (4095 muestras) se ejecuta el algoritmo de separación en sí. Cada parcial rastreado se caracteriza por su frecuencia promedio:

$$\overline{f_j}(t) = \frac{\sum_{i=1}^{N_T} f_{ins,j}(t_i)}{N_T} \tag{III.7}$$

donde N_T es el tamaño de trama ($N_T = 2^n - 1$, en este caso $n = 12$. Véase al respecto la Sección 3.9.3).

III.c.1. Armonicidad y distancia

En esta segunda aproximación, se utiliza una técnica detección de fundamentales muy básica para localizar las diferentes frecuencias base presentes en la señal. Dentro de un frame se busca el *mínimo* de las frecuencias características obtenidas a través de la Ecuación (III.7). Ésta será considerada como la fundamental de la primera fuente, f_{01}. A continuación

se buscan los armónicos de ésta, hf_{01} con h entero (con cierta tolerancia), siendo tales parciales considerados asimismo como miembros de la fuente s_1. Una vez barridos todos los parciales, de entre los que hayan permanecido como no asignados se busca el mínimo, que será f_{02} (fuente s_2) y el proceso se repite de forma recursiva. Al término del mismo, un mismo parcial puede haber sido asignado a más de una fuente.

Para asignar los parciales detectados a sus correspondientes fuentes, se emplea la *distancia armónica*, en la que se compara la frecuencia promedio del parcial, $\overline{f_j}$, con la frecuencia fundamental de cada fuente, f_{0k}, de acuerdo con la expresión:

$$d'_a(j,k) = \frac{f_{0k}}{\overline{f_j}}\, round\left(\frac{\overline{f_j}}{f_{0k}}\right) \tag{III.8}$$

Con esta definición, y teniendo en cuenta el método de detección empleado, la distancia armónica está acotada:

$$0{,}5 \leq d'_a(j,k) \leq 1{,}5 \quad \forall j,k \tag{III.9}$$

y $d'_a(j,k) \to 1 \Rightarrow P_j$ "es armónico de" s_k. En otras palabras, la decisión de si un parcial es o no armónico de una fuente determinada se tomará en función de un parámetro umbral, cuando:

$$|d'_a(j,k) - 1| < \theta_a \tag{III.10}$$

donde el valor del umbral tomado en la aplicación experimental del método es $\theta_a = 0.02$.

Los parciales que hayan sido asignados mediante esta distancia a una única fuente no ofrecen dudas: son los parciales aislados. Aquellos que hayan sido asignados a más de una (parciales superpuestos), serán reubicados utilizando para ello un nuevo patrón de similitud basado en esta ocasión en tres distancias: modular, frecuencial y armónica (por orden creciente de importancia). Dado que la distancia armónica $d'_a(j,k)$ de la Ecuación (III.8) no indica afinidad del mismo modo que las otras dos distancias, d_m y d_f, definidas de modo equivalente a las Ecuaciones (III.1) y (III.2) respectivamente, es necesario redefinirla de forma que la armonicidad de un parcial respecto de una fundamental dada quede marcada por distancias tendientes a 0 y la no-armonicidad por distancias tendientes a 1:

$$d_a(j,k) = \frac{|d'_a(j,k) - 1|}{d'_a(j,k)} \tag{III.11}$$

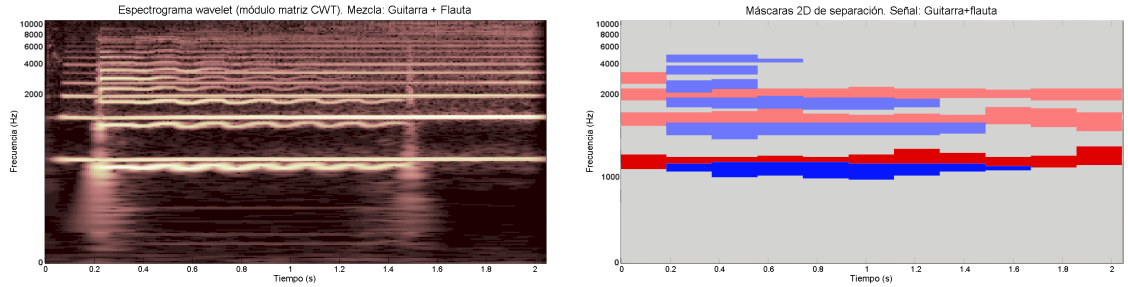
De este modo, la distancia total será:

$$d(i,j) = w_a d_a(i,j) + w_m d_m(i,j) + w_f d_f(i,j) \tag{III.12}$$

donde w_a =0.9, w_f=0.075 y w_m=0.025 son los pesos relativos de cada una de las tres distancias involucradas (cabe resaltar la gran importancia que recibe la distancia armónica

frente a las demás). Por su parte $d_m(i,j)$ y $d_f(i,j)$ están definidas, como se ha dicho, de forma equivalente a las Ecuaciones (III.1) y (III.2), con la salvedad de que se han normalizado previamente a 1.

El resultado de este proceso es, de nuevo, la generación de una máscara bidimensional para cada una de las fuentes detectadas en la mezcla. En la Figura III.10 se presentan los resultados para una señal mezcla de guitarra y flauta.



(a) *Espectrograma. Señal: mezcla de flauta y guitarra.* (b) *Máscaras. Señal: mezcla de flauta y saxo.*

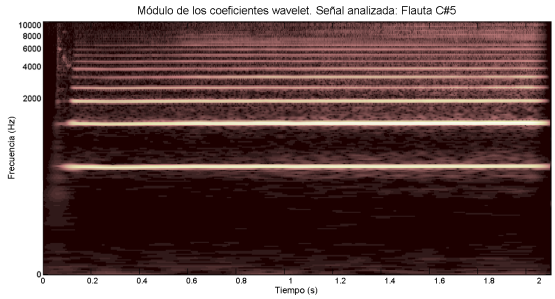
Figura III.10: *Espectrograma original y máscaras de separación de una señal mezcla de flauta y guitarra, obtenidas a partir del método de distancia armónica. (a) Espectrograma wavelet. (b) En azul, máscara correspondiente a la guitarra. En rojo, máscara de la flauta. En tonos más oscuros, los parciales fundamentales de cada fuente.*

En la Figura III.10(a) aparece el espectrograma wavelet de la señal mezcla de flauta y guitarra, mientras que en la Figura III.10(b) se han señalado las máscaras correspondientes a las dos fuentes detectadas (en rojo la guitarra, en azul la flauta, en tonos más oscuros los respectivos parciales fundamentales).

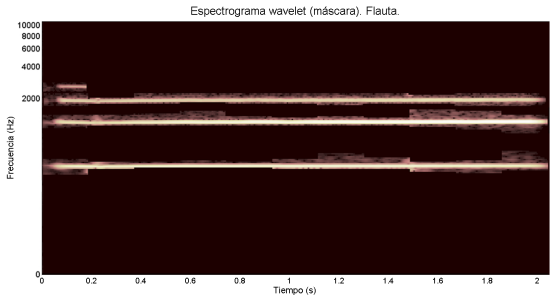
Por otro lado, en la Figura III.11 se han representado los espectrogramas de las fuentes separadas, comparados con los datos extraídos de la señal mezcla.

En las Figuras III.11(a) y III.11(b) aparecen respectivamente es espectrograma de la flauta aislada y el resultado de aplicar la máscara correspondiente a los coeficientes de la mezcla (Figura III.10). Por su parte, en las Figuras III.11(c) y III.11(d) se presentan el espectrograma de la guitarra (la señal original se ha extendido para que su duración coincida con la de la mezcla y los datos resulten visualmente comparables) y la aplicación de la máscara de la guitarra a los datos de la señal mezclada. Como se puede ver en las figuras de la derecha, la cantidad de información recogida *respecto al área total disponible* es baja. Sin embargo, en términos energéticos, se ha recuperado información suficiente[3], con

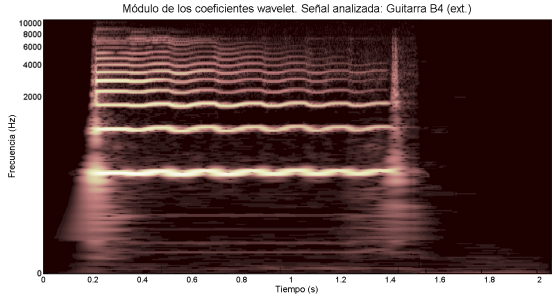
[3] *En el escalograma de la Figura 5.33(a) se puede apreciar que los tres primeros parciales de la flauta aportan la mayor parte de su energía. Por su parte, respecto a la guitarra, en la Sección 5.4.2.2 (en la representación tridimensional de la información) se presenta una prueba que apoya visualmente la misma*



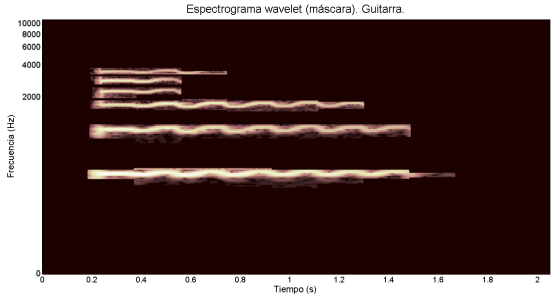
(a) *Espectrograma. Señal: Flauta C#5.*



(b) *Máscara. Flauta C#5 de la mezcla de guitarra y flauta.*



(c) *Espectrograma. Señal: Guitarra B4 extendida.*



(d) *Máscara. Guitarra B4 de la mezcla de guitarra y flauta.*

Figura III.11: *Parte izquierda: módulo de los coeficientes wavelet de las señales aisladas. Parte derecha: máscaras extraídas de la mezcla. (a) Espectrograma de la flauta C#5 aislada. (b) Máscara correspondiente a la flauta, extraída del espectrograma mezcla de guitarra y flauta. (c) Espectrograma de la guitarra B4 aislada (extendida). (d) Máscara correspondiente a la guitarra, extraída del espectrograma mezcla de guitarra y flauta.*

lo que los resultados sonoros son bastante fieles a las señales aisladas (con la citada y ahora evidente pérdida de *color* aducida con anterioridad).

III.c.2. Resultados, limitaciones y valoración

La técnica se ha puesto a prueba con un conjunto de señales de instrumentos musicales mezclados sintéticamente de dos en dos. Las señales utilizadas como base de las mezclas son: Un violín ejecutando una nota $G4$, un clarinete $F\#3$, una flauta $C\#5$ (sin vibrato), un saxo $C4$ (con vibrato) y una nota $B4$ de guitarra. De nuevo estas señales proceden de una pequeña base de datos propia. Los resultados numéricos de evaluación de calidad aparecen reflejados en las Tablas III.5 a III.7. Una vez más, el nombre de las señales está codificado

afirmación. Sin embargo, en señales en las que la energía se encuentre más repartida entre los armónicos, esta conclusión deja de ser cierta.

según se especifica en el Apéndice III.a.

Señal	SDR_1	SDR_2	$\overline{SDR}$
CF#3+FnvC#5	14.781	2.2009	8.4595
CF#3+GB4	22.2060	15.3780	18.7920
CF#3+SvC4	10.5412	2.1974	6.3693
CF#3+VG4	19.4883	16.1450	17.8166
FnvC#5+GB4	22.4208	21.8970	22.1589
FnvC#5+SvC4	12.9235	8.0454	10.4844
FnvC#5+VG4	13.3721	14.6615	14.0168
GB4+SvC4	17.7948	14.8484	16.3216
GB4+VG4	11.0684	11.8297	11.4491
SvC4+VG4	6.7342	16.7793	11.7568

Tabla III.5: *Separación por distancias. Resultados numéricos: Parámetro SDR.*

Señal	SIR_1	SIR_2	$\overline{SIR}$
CF#3+FnvC#5	19.9909	39.8751	29.9330
CF#3+GB4	37.3722	52.7652	45.0687
CF#3+SvC4	15.4965	46.4719	30.9842
CF#3+VG4	57.7680	75.4616	66.6148
FnvC#5+GB4	60.2400	50.2245	55.2323
FnvC#5+SvC4	22.3957	70.0924	46.2440
FnvC#5+VG4	38.3723	34.2106	36.2914
GB4+SvC4	31.0936	44.6581	37.8758
GB4+VG4	53.7400	29.2925	41.5163
SvC4+VG4	62.2650	32.2788	47.2719

Tabla III.6: *Separación por distancias. Resultados numéricos: Parámetro SIR (dB).*

Con estos resultados, los valores promediados de los parámetros para las señales analizadas son:

- $\overline{SDR_a}$= 13.7625 dB.
- $\overline{SIR_a}$= 43.7032 dB.
- $\overline{SAR_a}$= 13.9899 dB.

Señal	SAR_1	SAR_2	$\overline{SAR}$
CF#3+FnvC#5	16.2616	2.2021	9.2469
CF#3+GB4	22.3410	15.3788	18.8599
CF#3+SvC4	12.3337	2.1977	7.2657
CF#3+VG4	19.4889	16.1450	17.8170
FnvC#5+GB4	22.4215	21.9035	22.1625
FnvC#5+SvC4	13.4688	8.0454	10.7571
FnvC#5+VG4	13.3864	14.7116	14.0490
GB4+SvC4	18.0063	14.8531	16.4297
GB4+VG4	11.0687	11.9134	11.4910
SvC4+VG4	6.7342	16.9061	11.8201

Tabla III.7: *Separación por distancias. Resultados numéricos: Parámetro SAR (dB).*

Se han comparado una vez más las formas de onda y los espectros de las señales aisladas con las de las separadas por este método. En este caso se ha ido más allá de la comparación visual presenta con la técnica anterior, definiéndose un error relativo numérico temporal y otro frecuencial mediante las expresiones:

$$\varepsilon_{k,t} = \frac{\sum_{i=1}^{L} \bar{s}_k(t_i)^2}{\sum_{l=1}^{L} s_k(t_i)^2} \tag{III.13}$$

$$\varepsilon_{k,f} = \frac{\sum_{j=1}^{J} \hat{\bar{S}}_k(f_j)^2}{\sum_{j=1}^{J} \hat{S}_k(f_j)^2} \tag{III.14}$$

donde $\bar{s}_k$ es la fuente $k-sima$ separada, s_k la señal aislada correspondiente y $\hat{\bar{S}}_k$, $\hat{S}_k$ sus respectivas transformadas de Fourier, mientras L es el número total de muestras de $x(t)$ y J la longitud espectral (número de escalas) de las señales. Los resultados numéricos obtenidos (en dB) aparecen presentados en la Tabla III.8.

De este segundo método de separación propuesto cabe destacar que la importancia de la distancia frecuencial y modular es mucho menor que en método de separación por onsets. La asignación de fuentes se hace, básicamente, atendiendo a criterios armónicos y los resultados sonoros obtenidos presentan de este modo una menor interferencia entre fuentes (lo que debe suponer una mejora en los parámetros numéricos de calidad de la separación). Esta tendencia se llevará al extremo en la tercera y última técnica, que será presentada a continuación.

El intento de separación de fuentes mediante máscaras bidimensionales con intersección nula queda limitado por la resolución frecuencial, y no pueden llegar mucho más allá de lo presentado en esta Sección. La superposición de parciales (inevitable en la zona de alta

Señal	$\varepsilon_{1,t}$ (dB)	$\varepsilon_{1,f}$ (dB)	$\varepsilon_{2,t}$ (dB)	$\varepsilon_{2,f}$ (dB)
CF#3+FnvC#5	-8.50	-10.33	-29.47	-44.85
CF#3+GB4	-30.90	-37.70	-44.46	-58.62
CF#3+SvC4	-8.49	-10.40	-21.20	-32.31
CF#3+VG4	-32.50	-36.35	-39.07	-50.09
FnvC#5+GB4	-44.89	-46.05	-43.85	-56.07
FnvC#5+SvC4	-25.93	-31.98	-17.36	-18.70
FnvC#5+VG4	-27.08	-27.19	-30.08	-32.83
GB4+SvC4	-35.71	-40.73	-29.93	-33.54
GB4+VG4	-22.79	-27.06	-24.19	-36.37
SvC4+VG4	-15.14	-17.26	-33.70	-46.48

Tabla III.8: *Separación por distancias. Resultados numéricos: Errores en tiempo y frecuencia (en dB).*

frecuencia) se da por definición en piezas musicales complejas, donde aparecen con frecuencia relaciones armónicas entre las fuentes mezcladas. Por lo tanto, un método de separación generalista debe incidir en el complicado caso de la separación de este tipo de información compartida, donde se abandona la idea de máscaras bidimensionales para tratar la información individual de cada parcial.

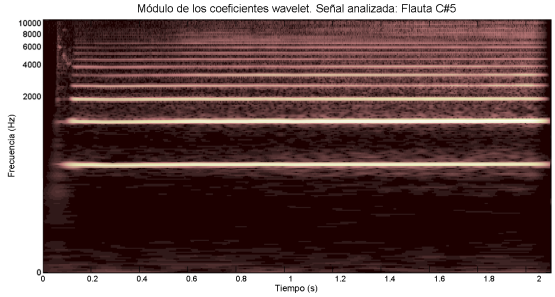
III.c.3. Resultados gráficos adicionales

A continuación se expondrán algunos resultados adicionales correspondientes al mismo ejemplo detallado en la Sección anterior, concretamente la mezcla de una guitarra $B4$ y una flauta $C\#5$.

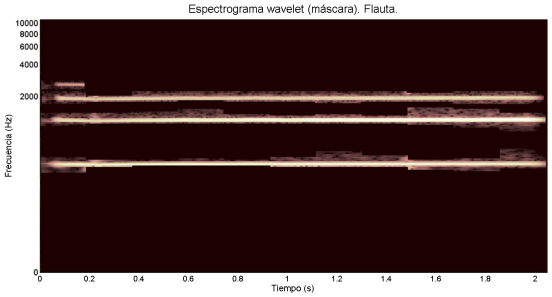
Tomamos como punto de partida las máscaras bidimensionales calculadas para cada fuente (Figura III.11), que se reproducen de nuevo en la Figura III.12. La conclusión aparente que se extrae de estas figuras es que la información contenida en las máscaras es escasa. Sin embargo, energéticamente esto no es cierto.

En la Figura III.13 se han representado los espectros correspondientes a cada una de las fuentes separadas, comparados con los de las señales aisladas. Como se puede ver, la mayor parte de la energía ha sido correctamente separada. Esta situación tiende a ser general en notas de fundamentales no superpuestas (y de relación no armónica) situadas en las octavas 3 y 4, y más cuestionable en octavas superiores e inferiores, en ambos casos debido a la excesiva superposición de tonos.

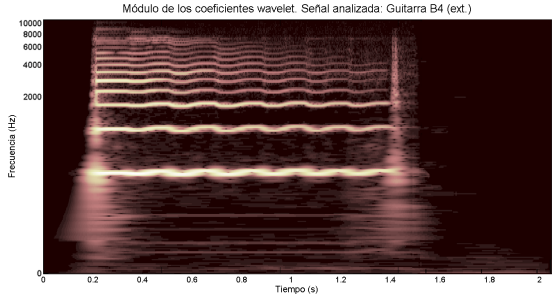
Como se puede apreciar en las máscaras y los espectros, los parciales de alta frecuencia (con niveles de superposición muy elevados) no son tenidos en cuenta en este método, lo



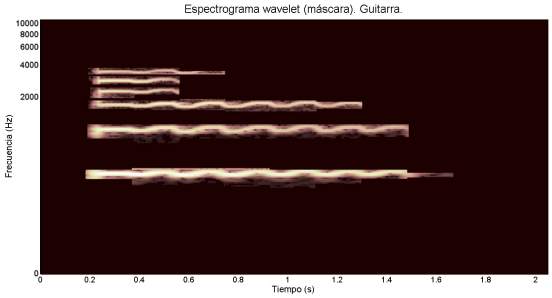
(a) *Espectrograma. Señal: Flauta C#5.*



(b) *Máscara. Flauta C#5 de la mezcla de guitarra y flauta.*



(c) *Espectrograma. Señal: Guitarra B4 extendida.*

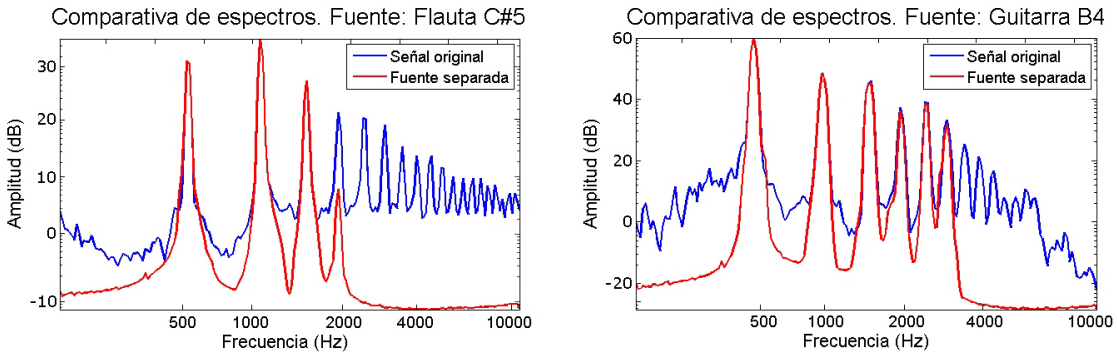


(d) *Máscara. Guitarra B4 de la mezcla de guitarra y flauta.*

Figura III.12: *Parte izquierda: módulo de los coeficientes wavelet de las señales aisladas. Parte derecha: máscaras extraídas de la mezcla. (a) Espectrograma de la flauta C#5 aislada. (b) Máscara correspondiente a la flauta, extraída del espectrograma mezcla de guitarra y flauta. (c) Espectrograma de la guitarra B4 aislada (extendida). (d) Máscara correspondiente a la guitarra, extraída del espectrograma mezcla de guitarra y flauta.*

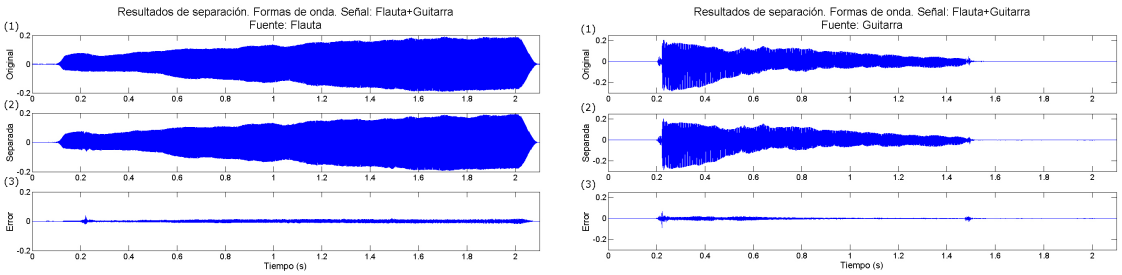
cual produce sonidos menos coloridos que los originales. Esta situación podría superarse reconstruyendo la parte armónica de la señal partiendo de los parciales así separados, lo cual nos llevaría a un método muy similar al presentado en la Sección 4.7, que será ampliado en el Apéndice III.e.

Para terminar, en la Figura III.14 se han representado las formas de onda correspondientes a las señales aisladas (flauta, Figura III.14(a) y guitarra, Figura III.14(b) respectivamente) y sus correspondientes fuentes separadas obtenidas por distancia armónica. Los errores tienden a ser menores que los obtenidos por el método de los onsets.



(a) *Flauta C#5: Espectros.* (b) *Guitarra B4: Espectros.*

Figura III.13: *Resultados de separación por el método de armonía. Espectros correspondientes a los resultados de la separación de la señal mezcla de guitarra B4 y flauta C#5. (a) En azul, espectro de la flauta original. En rojo, espectro de la flauta separada (eje vertical en dB). (b) En azul, espectro de la guitarra original. En rojo, espectro de la guitarra separada (eje vertical en dB).*



(a) *Fuente: Flauta.* (b) *Fuente: Guitarra.*

Figura III.14: *Resultados de separación por el método de distancia armónica. Formas de onda correspondientes a los resultados de la separación de la señal mezcla de guitarra B4 más flauta C#5. (a) (1) Flauta original. (2) Flauta separada. (3) Error. (b) (1) Guitarra original. (2) Guitarra separada. (3) Error.*

III.d. Parciales superpuestos: análisis de fase

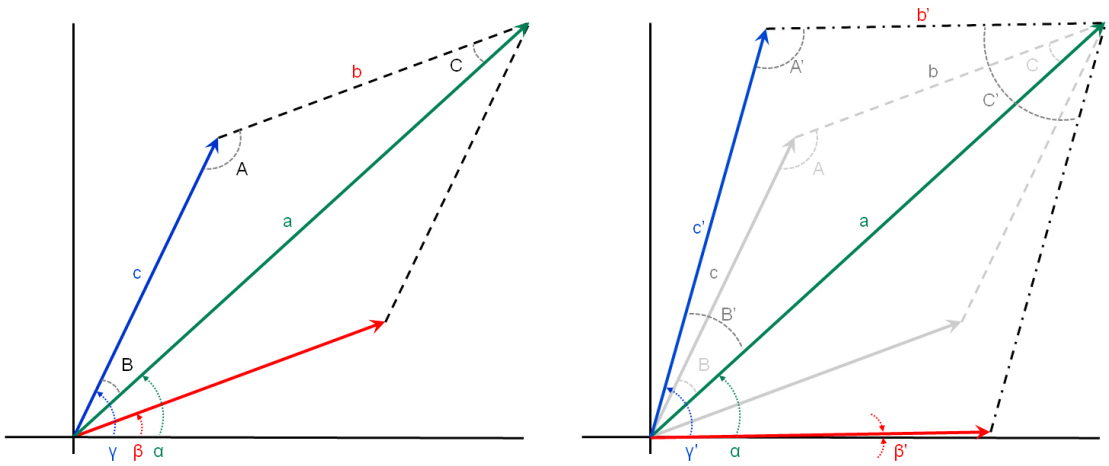
En la Sección 4.7.2 se afirma que, incluso conociendo las amplitudes teóricas de los parciales a reconstruir, un error de una parte entre 10^3 en la detección de la fase inicial desvirtúa rápidamente los resultados. Este resultado se ha alcanzado tras analizar matemáticamente la suma de parciales complejos vista como suma vectorial.

Por simplicidad, se va a resolver el caso de dos parciales superpuestos. Sean $u(t) =$

$b(t)e^{j\beta(t)}$ y $v(t) = c(t)e^{j\gamma(t)}$ dos números complejos cualesquiera (en este caso, los correspondientes a los datos del parcial eventualmente superpuesto, aunque la situación matemática es mucho más general). Su suma $w(t)$ puede escribirse como:

$$w(t) = a(t)e^{j\alpha(t)} = b(t)e^{j\beta(t)} + c(t)e^{j\gamma(t)} \tag{III.15}$$

Evaluada en un instante de tiempo cualquiera, la situación codificada en la Ecuación (III.15) queda plasmada en la Figura III.15(a).



(a) *Suma original.* (b) *Suma estimada (indeterminación).*

Figura III.15: *Carácter indeterminado de la suma vectorial de dos números complejos u y v que hace necesaria la estimación de parámetros para resolver el problema. (a) Suma vectorial de dos números complejos $u = be^{j\beta}$ (en rojo) y $v = ce^{j\gamma}$ (en azul) para dar $w = ae^{j\alpha}$ (en verde). (b) El mismo vector suma $w = ae^{j\alpha}$ puede ponerse como combinación de infinitos pares de $u = be^{j\beta}$ y $v = ce^{j\gamma}$.*

Basándonos en la combinación lineal de la Figura III.15(a), se pueden realizar una serie de cálculos básicos muy simples para obtener el valor de las amplitudes de los vectores u y v en función de las fases involucradas. Aplicando la ley de los senos [159] al triángulo plano marcado por los vértices de ángulos A, B y C, se tiene:

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C} \tag{III.16}$$

donde:

$$A = \pi - \gamma + \beta \tag{III.17}$$

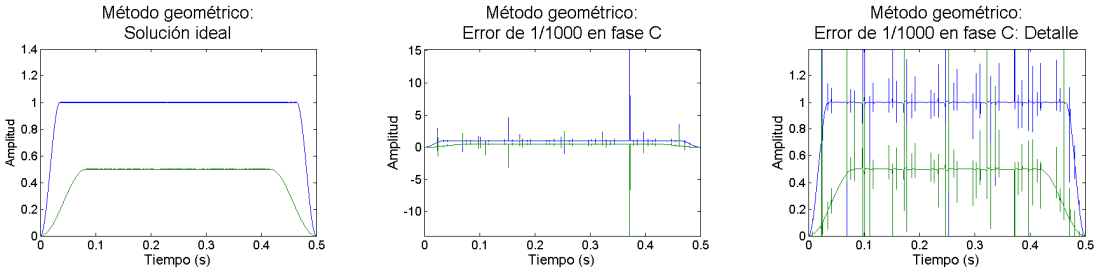
$$B = \gamma - \alpha \tag{III.18}$$

$$C = \alpha - \beta \tag{III.19}$$

En la separación, el único dato al que se tiene acceso es el valor tanto de amplitud como de fase del parcial mezcla, sean a y α por simplicidad. Es evidente (Figura III.15(b)) que la información disponible resulta insuficiente para resolver adecuadamente el problema: la suma no es una función biunívoca. Existen infinitos pares de escalares o vectores que, sumados, arrojan es el mismo resultado. En la Figura III.15(b) se han reflejado dos de estos pares, determinados en coordenadas polares por (b, β) y (c, γ), por un lado (en gris, correspondientes con los datos de la Figura III.15(a) y supuestos como los datos correctos de la separación) y (b', β') junto con (c', γ'), que representan otra de las infinitas parejas cuya suma es (a, α).

Desde el instante en que no existe una única solución para la suma, se entra en el terreno de la estimación: para llegar a la solución correcta, se deben hacer caracterizar u y v con un grado de aproximación suficiente. El problema se reduce, evidentemente, al diseño de un algoritmo de estimación de la fase inicial. Sin embargo, la precisión en esta estimación debe ser extrema (y las herramientas de verificación de resultados no están muy claras, sobre todo para el caso de señales reales).

Se ha diseñado un pequeño experimento en el cual se mezclan dos señales de amplitudes (con envolventes de Hanning) $A_1 = 1$ y $A_2 = 0.5$ y frecuencias $f_1 = 1000$Hz y $f_2 = 1033$Hz. Los resultados del mismo aparecen en la Figura III.16.



(a) *Resultado ideal.* (b) *Con desfase de* 10^{-3}. (c) *Detalle.*

Figura III.16: *Resultados de separación aplicando el método geométrico de los tres senos. (a) Conociendo las fases involucradas. (b) Resultado insertando un desfase de 1/1000 en el ángulo C. (c) Detalle de las amplitudes obtenidas con el desfase de 1/1000.*

Aplicando directamente el método geométrico de la ley de los senos descrito con anterioridad, y conociendo todas las fases involucradas, las amplitudes estimadas aparecen en la Figura III.16(a). Sin embargo, añadiendo un pequeño error de 1/1000 en la fase C, los resultados quedan claramente distorsionados, como se refleja en las amplitudes representadas en las Figuras III.16(c) y III.16(c), donde se pueden observar grandes discontinuidades en

los valores de las envolventes. La causa de estos saltos es evidente. De la Ecuación (III.16) se infiere que las tres sinusoides involucradas deben anularse simultáneamente. En cuanto cualquiera de ellas no se comporte de este modo, se producen divergencias en los resultados de las amplitudes involucradas. Y eso es exactamente lo que sucede al insertar ese pequeño desfase en C. Errores menores en la estimación de la fase provocan saltos de amplitud decreciente. Con un error aproximado de una parte por 10^5, los errores resultan asumibles.

La dificultad en la estimación de las fases involucradas en una mezcla con tal precisión es poco menos que imposible (o excesivamente costoso en tiempo de computación), lo cual inhabilita este método como estimación de los parciales superpuestos y nos lleva a emplear técnicas de estimación energética, como la empleada en la Sección 4.7.

III.e. Parciales superpuestos: detalles

A continuación se ampliarán los resultados experimentales en la medida de calidad de separación. Estos resultados adicionales se corresponden con datos numéricos concretos de algunos de los experimentos realizados, así como detalles en la obtención de las señales que asisten en la obtención de los parámetros numéricos de calidad empleados.

III.e.1. Datos numéricos

En primer lugar se van a incluir los resultados numéricos detallados correspondientes a 3 de los 8 ensayos llevados a cabo, en concreto mezclas de 2 y 3 instrumentos ejecutando notas aleatorias y los resultados obtenidos en la separación de un instrumento armónico mezclado con uno inarmónico (piano).

Los parámetros de calidad SDR, SIR y SAR aparecen en las Tablas III.9 a III.11. En éstas se incluyen los resultados experimentales obtenidos para cada una de las fuentes encontradas y separadas así como el promedio de los mismos (un indicativo de la calidad promedio de la separación de cada mezcla).

III.e.2. Resultados gráficos adicionales

Si bien se dispone de los datos explícitos de cada uno de los espectrogramas y escalogramas wavelet, formas de onda y espectros de todas las señales separadas por reconstrucción de amplitud y fase, añadir todos estos resultados sería excesivo incluso para un Anexo. Sin embargo, como apoyo a la robustez del método propuesto y a los datos presentados en la Sección 4.7, a continuación se incluyen las gráficas correspondientes a dos señales mezcla de dos fuentes y otras dos de tres.

En la Figura III.17, los espectrogramas y escalogramas wavelet de las 4 señales. El valor elevado de D se deduce de la mayor precisión en la localización frecuencial (compárense

Mezcla de dos fuentes armónicas			
Señal	SDR_1	SDR_2	$\overline{SDR}$
FnvC#5+GB4	30.9780	15.0499	23.0139
BD4+HG4	22.1792	12.2720	17.2256
TrTC5+TvD5	32.9437	28.1849	30.5643
VG4+GB4	16.4476	16.5802	16.5139
CeC#4+FvB4	26.2550	17.3567	21.8059
TnvC5+CbC#4	30.8267	30.8069	30.8168
SC3+FnvC#5	19.1246	24.9896	22.0571
TnvB5+OF5	25.3510	33.8416	29.5963
FvB3+SSG#3	16.6133	17.9595	17.2864
HF#4+BD#4	29.5651	29.0708	29.3180
AFA#5+OF#5	27.8945	18.0306	22.9625
TuC4+ViA#3	25.2563	27.1284	26.1924
BCD2+BFC3	12.5086	17.3728	14.9407
AFE4+AFF#5	27.1284	25.2563	26.1924
VA#4+CeG#4	20.3137	23.0616	21.6876

Mezcla de dos fuentes (un instrumento inarmónico)			
Señal	SDR_1	SDR_2	$\overline{SDR}$
TnvB5+PG#3	24.6972	10.3093	17.5033
CbD#4+PG#3	20.7072	-3.3710	8.6681
FvB3+PG#3	13.9715	-7.1017	3.4349

Mezcla de tres fuentes armónicas				
Señal	SDR_1	SDR_2	SDR_3	$\overline{SDR}$
ASC#4+BE4+GB4	20.3099	21.7320	6.3413	16.1278
TvC5+FnvA4+OD5	22.1084	12.1312	13.1094	15.7830
TrBF#3+SSvG#3+FvC#4	18.1797	18.9864	11.5468	16.2376
HC4+VA4+CeD4	22.7484	2.9655	2.4187	9.3775
TnvE5+CbD5+TrTC5	19.6558	35.5402	16.2889	23.8283

Tabla III.9: *Método de separación de parciales superpuestos: Resultados numéricos del parámetro SDR (dB).*

estas gráficas con, por ejemplo, cualquiera de los espectrogramas presentados en la Figura III.7).

En la Figura III.18, las formas de onda obtenidas en el proceso de separación de las mezclas de dos fuentes. En la Figura III.19, las correspondientes de las mezclas de tres fuentes.

Mezcla de dos fuentes armónicas			
Señal	SIR_1	SIR_2	$\overline{SIR}$
FnvC#5+GB4	64.0949	86.3455	75.2202
BD4+HG4	72.4347	60.1982	66.3165
TrTC5+TvD5	90.5903	84.8376	87.7139
VG4+GB4	55.0862	50.9414	53.0138
CeC#4+FvB4	110.5218	78.9238	94.7228
TnvC5+CbC#4	76.1676	66.9925	71.5801
SC3+FnvC#5	64.7522	58.5935	61.6729
TnvB5+OF5	85.6674	74.6080	80.1377
FvB3+SSG#3	71.0086	80.8125	75.9105
HF#4+BD#4	65.9539	71.7516	68.8527
AFA#5+OF#5	70.7206	68.1027	69.4117
TuC4+ViA#3	82.3606	64.5388	73.4497
BCD2+BFC3	62.1065	55.8376	58.9720
AFE4+AFF#5	64.5388	82.3606	73.4497
VA#4+CeG#4	66.2725	66.7609	66.5167

Mezcla de dos fuentes (un instrumento inarmónico)			
Señal	SIR_1	SIR_2	$\overline{SIR}$
TnvB5+PG#3	66.8617	67.1622	67.0120
CbD#4+PG#3	50.5429	72.4518	61.4974
FvB3+PG#3	22.9468	37.2872	30.1170

Mezcla de tres fuentes armónicas				
Señal	SIR_1	SIR_2	SIR_3	$\overline{SIR}$
ASC#4+BE4+GB4	41.7025	51.0600	60.3403	51.0343
TvC5+FnvA4+OD5	39.4132	45.7920	53.6480	46.2844
TrBF#3+SSvG#3+FvC#4	67.8138	69.7338	67.6455	68.3977
HC4+VA4+CeD4	58.3956	41.2671	32.0106	43.8911
TnvE5+CbD5+TrTC5	81.3426	71.7489	69.1010	74.0642

Tabla III.10: *Método de separación de parciales superpuestos: Resultados numéricos del parámetro SIR (dB).*

A falta del auténtico test de calidad, no ya basado en el valor de los parámetros estándar, sino en la semejanza sonora entre las señales involucradas, las formas de onda de estas figuras son un indicativo indirecto de la gran semejanza entre las señales originales y las fuentes separadas. Los datos numéricos que respaldan esta afirmación están codificados en las Tablas III.5 a III.11 de la Sección III.e.1. Pero, una vez más, la mejor forma de comprobar

Mezcla de dos fuentes armónicas			
Señal	SAR_1	SAR_2	$\overline{SAR}$
FnvC#5+GB4	30.9801	15.0499	23.0150
BD4+HG4	22.1792	12.2721	17.2257
TrTC5+TvD5	32.9437	28.1849	30.5643
VG4+GB4	16.4483	16.5818	16.5150
CeC#4+FvB4	26.2550	17.3567	21.8059
TnvC5+CbC#4	30.8269	30.8080	30.8174
SC3+FnvC#5	19.1247	24.9915	22.0581
TnvB5+OF5	25.3510	33.8420	29.5965
FvB3+SSG#3	16.6133	17.9595	17.2864
HF#4+BD#4	29.5661	29.0710	29.3186
AFA#5+OF#5	27.8947	18.0306	22.9627
TuC4+ViA#3	25.2563	27.1292	26.1928
BCD2+BFC3	12.5087	17.3734	14.9410
AFE4+AFF#5	27.1292	25.2563	26.1928
VA#4+CeG#4	20.3138	23.0617	21.6878

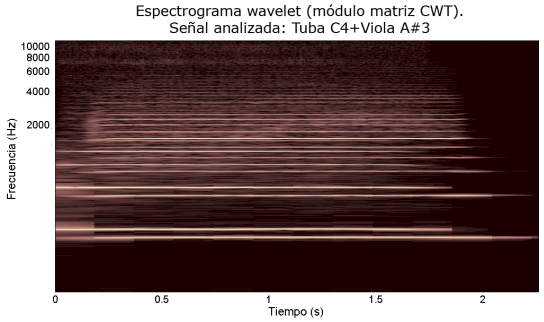
Mezcla de dos fuentes (un instrumento inarmónico)			
Señal	SAR_1	SAR_2	$\overline{SAR}$
TnvB5+PG#3	24.6974	10.3093	17.5034
CbD#4+PG#3	20.7117	-3.3710	8.6704
FvB3+PG#3	14.5814	-7.1008	3.7403

Mezcla de tres fuentes armónicas				
Señal	SAR_1	SAR_2	SAR_3	$\overline{SAR}$
ASC#4+BE4+GB4	20.3418	21.7371	6.3413	16.1401
TvC5+FnvA4+OD5	22.1905	12.1332	13.1098	15.8112
TrBF#3+SSvG#3+FvC#4	18.1798	18.9864	11.5468	16.2377
HC4+VA4+CeD4	22.7496	2.9664	2.4262	9.3807
TnvE5+CbD5+TrTC5	19.6558	35.5412	16.2889	23.8287

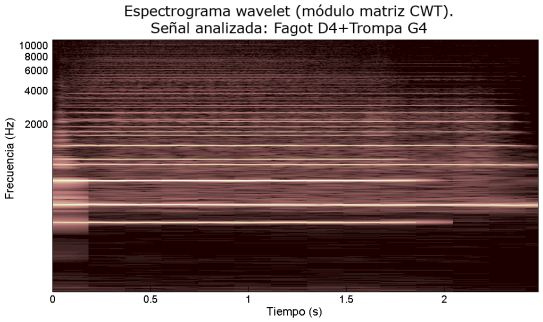
Tabla III.11: *Método de separación de parciales superpuestos: Resultados numéricos del parámetro SAR (dB).*

la calidad de la separación es escuchar los sonidos correspondientes, presentes en el soporte digital adjunto al presente documento.

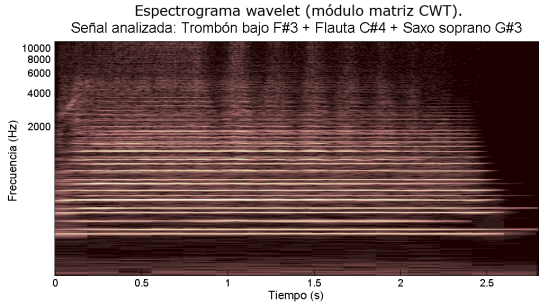
El algoritmo propuesto es capaz de reconstruir de forma bastante aproximada la parte armónica de las fuentes mezcladas. Para constatar este hecho, se puede recurrir una vez más a los espectros de las señales. En la Figura III.20, los espectros de Fourier correspondientes



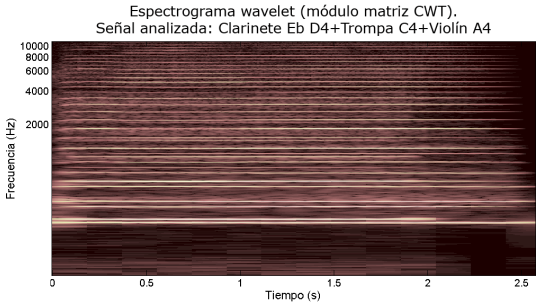
(a) *Espectrograma. Señal: mezcla de tuba y viola.*



(b) *Espectrograma. Señal: mezcla de fagot y trompa.*



(c) *Espectrograma. Señal: mezcla de trombón bajo, flauta y saxo soprano.*

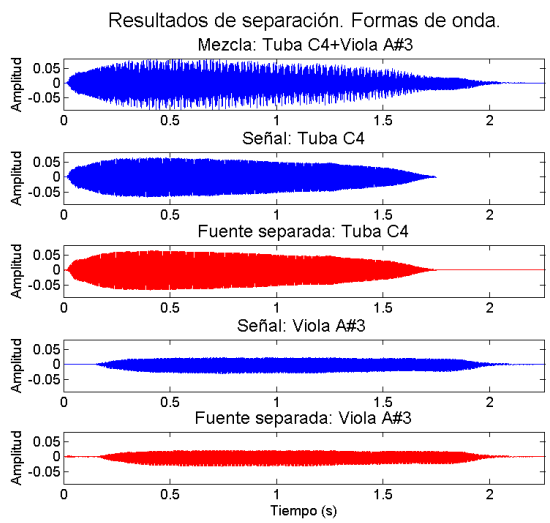


(d) *Espectrograma. Señal: mezcla de trompa, clarinete en mi bemol y violín.*

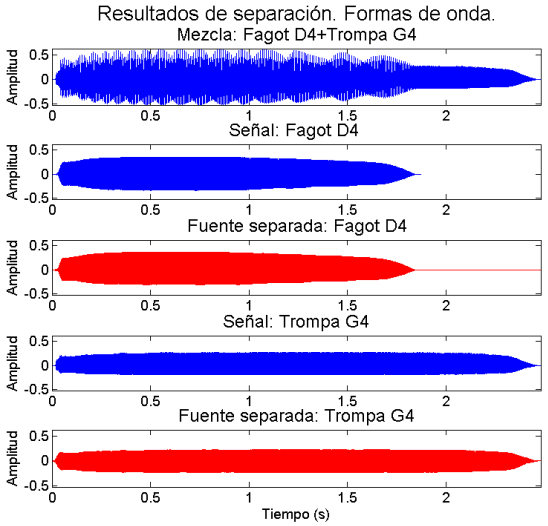
Figura III.17: *Módulo de los coeficientes wavelet (espectrograma wavelet) de cuatro de las señales analizadas. (a) Señal mezcla de tuba y viola. (b) Señal mezcla de fagot y trompa. (c) Señal mezcla de trombón bajo, flauta y saxo soprano. (d) Señal mezcla de trompa, clarinete en mi bemol y violín.*

a estas 4 señales de dos y tres fuentes. En azul, como siempre, los espectros de las señales aisladas. En rojo, los de las fuentes separadas. En la gráfica superior de cada subfigura, el espectro correspondiente a la señal mezcla.

Para finalizar, en las Ecuaciones (4.1), (4.3) y (4.4) se caracterizan respectivamente las señales de distorsión, interferencia y artefactos, partiendo de las definiciones propuestas en [66, 71, 165]. Estas señales están relacionadas, a su vez, con las demás fuentes presentes en la mezcla así como con espurios inherentes al proceso de separación. Estas señales son un indicativo más acerca de la calidad de la separación, ya que cuanto menor sea su energía (valor RMS), tanto mejor será el correspondiente coeficiente estándar de calidad. En la Figuras III.21 y III.22 se incluye un ejemplo de las formas de onda a partir de las que se obtienen estas señales, respectivamente para los casos de la mezcla de dos fuentes (tuba y viola) y de tres fuentes (trombón bajo, flauta y saxo soprano).



(a) *Formas de onda. Señal: mezcla de tuba y viola.*

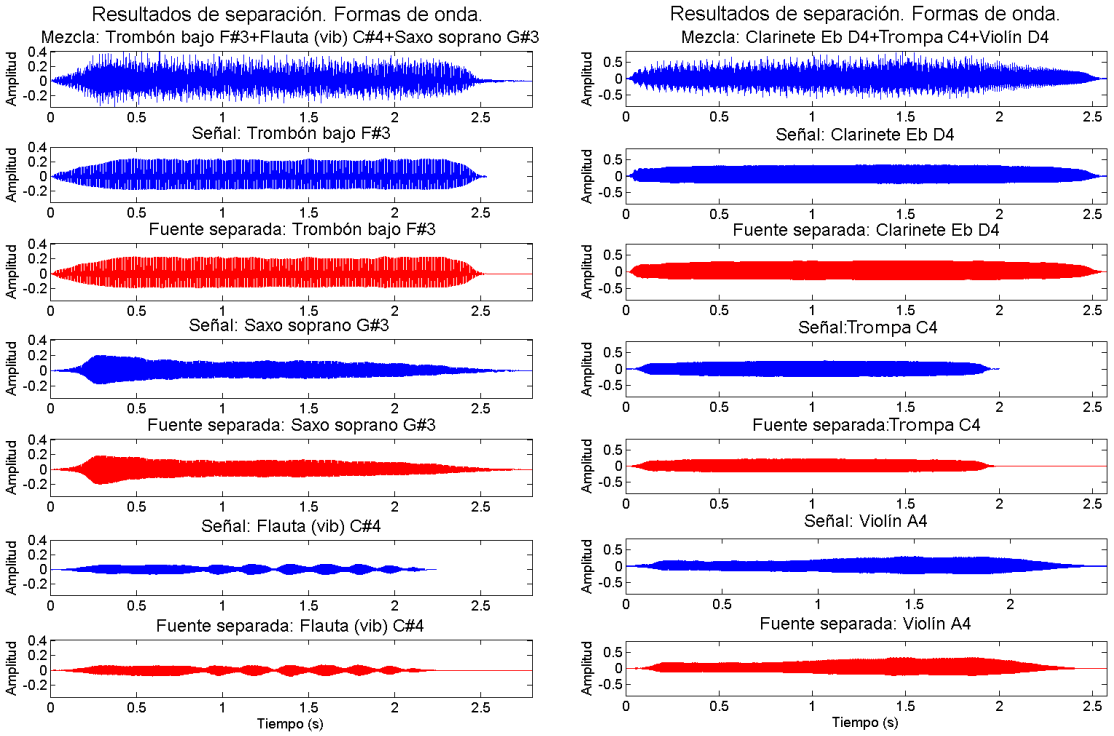


(b) *Formas de onda. Señal: mezcla de fagot y trompa.*

Figura III.18: *Resultados de la separación. Gráfica superior (de ambas subfiguras): señal mezcla. Debajo, en azul: señales originales. En rojo: fuentes separadas. (a) Mezcla de tuba y viola. (b) Mezcla de fagot y trompa.*

La señal objetivo es la señal aislada; las señales de interferencia y de artificios se obtienen utilizando ésta señal objetivo y la correspondiente fuente separada, y siguiendo las Ecuaciones (4.1), (4.3) y (4.4), como se ha explicado en el párrafo anterior. Del orden de magnitud de las señales de error se infiere el valor en dB de los correspondientes parámetros estándar de calidad, SDR, SIR y SAR, que vienen a ser un promedio energético de estas señales, expresado en dB (Sección 4.3).

A la hora de publicar el presente documento, la cantidad y variedad de experimentos de separación de fuentes ha sido aumentada considerablemente. En cuanto a la separación de dos fuentes, se han realizado nuevas pruebas que incluyen la separación de un mismo instrumento, tanto ejecutando dos notas diferentes dentro de la misma octava, como ejecutando notas armónicamente relacionadas (concretamente $C-G$, $D-A$, $E-B$, $F-C$, $G-D$, $A-E$ y $A\#-F$). Éstas mismas relaciones armónicas, pero ejecutadas por instrumentos diferentes, han sido asimismo analizadas y separadas. Respecto a las mezclas de tres fuentes, se han realizado dos nuevos experimentos: la separación de instrumentos ejecutando un acorde mayor ($C-E-G$) y un acorde menor ($A-C-E$). El total de señales analizadas hasta el momento ronda el medio centenar. Considerando globalmente los experimentos sobre mezcla de dos fuentes, los resultados promedio de medida en la calidad de la separación



(a) *Formas de onda. Señal: mezcla de trombón bajo, flauta y saxo soprano.*

(b) *Formas de onda. Señal: mezcla de trompa, clarinete en mi bemol y violín.*

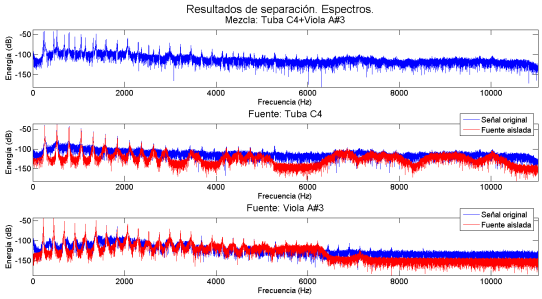
Figura III.19: *Resultados de la separación. Gráfica superior (de ambas subfiguras): señal mezcla. Debajo, en azul: señales originales. En rojo: fuentes separadas. (a) Mezcla de trombón bajo, flauta y saxo soprano. (b) Mezcla de trompa, clarinete en mi bemol y violín.*

son:

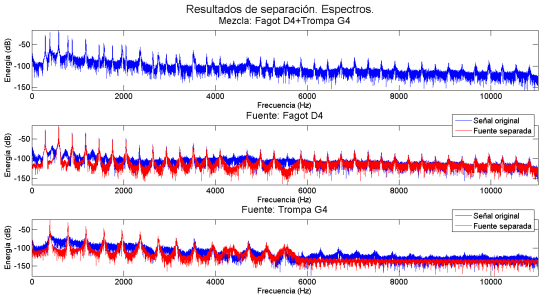
- $\overline{SDR_{2s}} \approx 17.77$ dB.
- $\overline{SIR_{2s}} \approx 60.61$ dB.
- $\overline{SAR_{2s}} \approx 17.78$ dB.

Y en cuanto a los resultados de los experimentos con tres fuentes, los parámetros estándar promedio quedan:

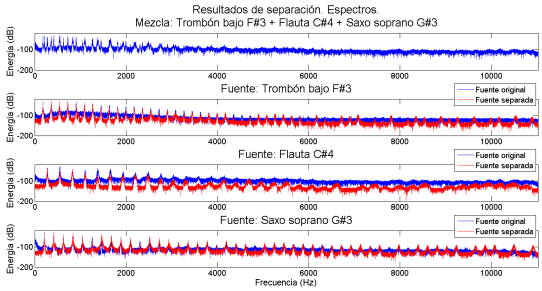
- $\overline{SDR_{3s}} \approx 13.78$ dB.
- $\overline{SIR_{3s}} \approx 53.99$ dB.



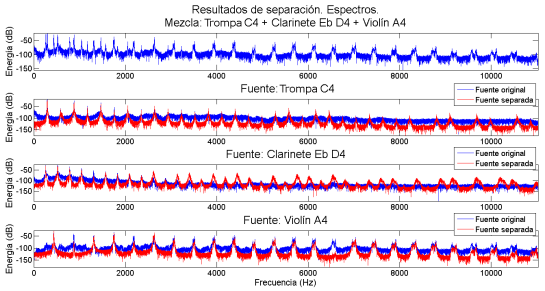
(a) *Espectros. Señal: mezcla de tuba y viola.*



(b) *Espectros. Señal: mezcla de fagot y trompa.*



(c) *Espectros. Señal: mezcla de trombón bajo, flauta y saxo soprano.*



(d) *Espectros. Señal: mezcla de trompa, clarinete en mi bemol y violín.*

Figura III.20: *Espectros de Fourier de cuatro de las señales analizadas. Gráfica superior: Espectro de la mezcla. Gráficas inferiores: espectros de las señales originales (en azul) y de las fuentes separadas (en rojo). (a) Señal mezcla de tuba y viola. (b) Señal mezcla de fagot y trompa. (c) Señal mezcla de trombón bajo, flauta y saxo soprano. (d) Señal mezcla de trompa, clarinete en mi bemol y violín.*

- $\overline{SAR_{3s}} \approx 13.79$ dB.

Realizando esta batería de experimentos de medida de calidad de la separación se ha podido constatar un hecho interesante, sobre el que se continuará trabajando: el valor de los coeficientes SDR, SIR y SAR no está ligado necesariamente a la calidad sonora final de la señal. Es decir, un valor elevado de estos coeficientes no es garantía absoluta de que la señal original y la fuente separada suenen de forma similar, y del mismo modo, un valor bajo de los mismos tampoco supone que el sonido separado tengan características audibles muy diferentes. Esto significa que se hace necesario encontrar en la literatura, o generar, si es necesario, un nuevo parámetro de calidad que mida de forma más realista las similitudes sonoras entre las señales aisladas y las fuentes separadas.

El modelo empleado para la estimación de parciales solapados emplea exclusivamente la armonía de los instrumentos musicales. Esto significa que instrumentos no armónicos, como el piano, no son adecuadamente reconstruidos en frecuencia (si bien esto no supone

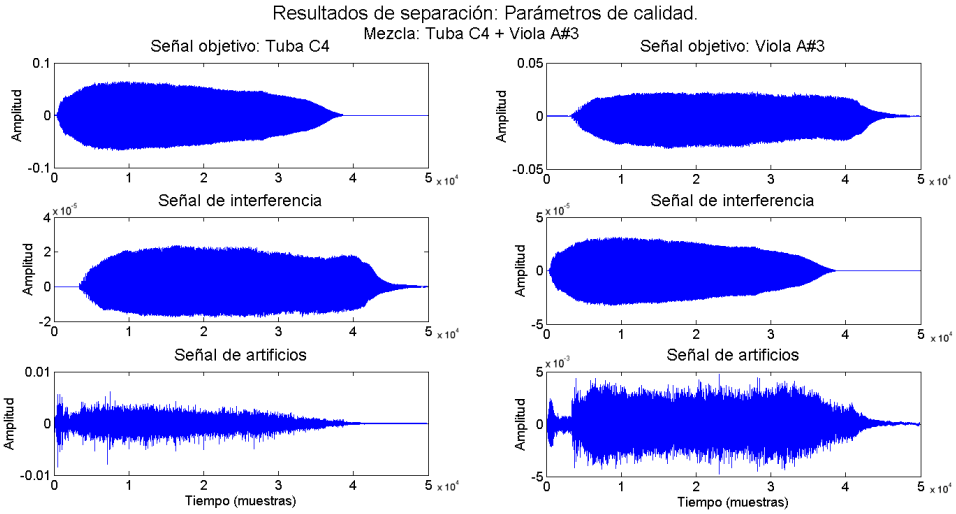


Figura III.21: *Resultados de la separación. Señales intermedias de calidad. Señales objetivo, patrones de interferencia y señales de artificios para la mezcla de tuba y viola.*

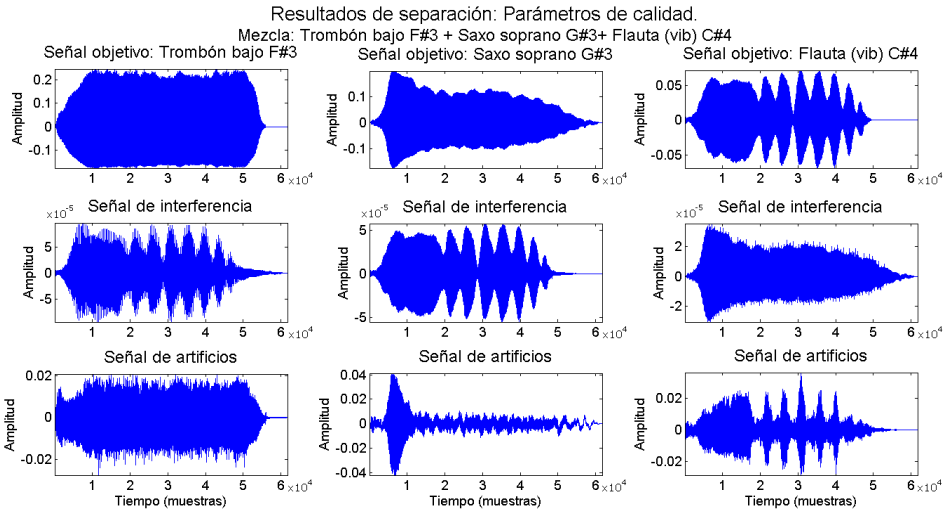


Figura III.22: *Resultados de la separación. Señales intermedias de calidad. Señales objetivo, patrones de interferencia y señales de artificios para la mezcla de trombón bajo, flauta y saxo soprano.*

necesariamente un problema acústico importante). La inclusión de este tipo de instrumentos no armónicos en el modelo parece un paso muy interesante de cara a generalizar sus

aplicaciones.

Cierto tipo de instrumentos (por ejemplo los instrumentos de viento), quedan muy caracterizados por información ruidosa subyacente (como el soplido del ejecutante). La técnica propuesta no incluye el tratamiento de parciales ruidosos o inarmónicos. El reparto de ésta energía no asignada supondría un aumento innegable de la calidad acústica de las señales separadas.

III.f. Valoración global de las técnicas de separación

La cantidad de señales analizadas en cada una de las tres técnicas de separación que se han desarrollado ha ido aumentando paulatinamente, si bien no es hasta el último caso cuando la solidez del método ha alcanzado un nivel satisfactorio. La comparativa de resultados entre estos tres algoritmos no puede ser completa hasta que la calidad sonora de las señales separadas es comprobada acústicamente. Sin embargo, y pese al riesgo de parecer sesgar favorablemente las conclusiones, se presentará una comparativa de los resultados numéricos que demuestra la evolución de la técnica.

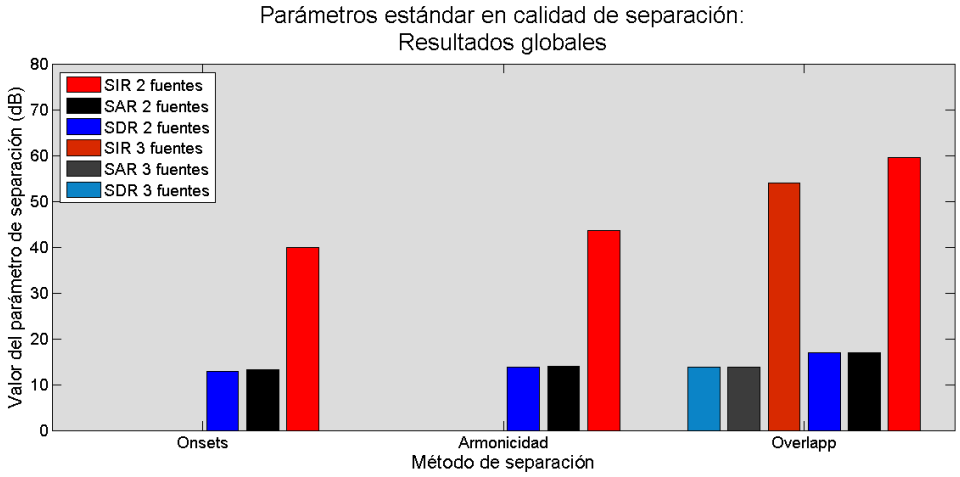


Figura III.23: *Comparativa de los resultados numéricos promedio obtenidos mediante cada una de las tres técnicas propuestas.*

En la Figura III.f se muestran los resultados numéricos promedio de los tres parámetros estándar para cada uno de los métodos presentados. En la gráfica se ha representado el diagrama de barras con los datos correspondientes a la separación de 2 fuentes $\overline{SDR_o}$, $\overline{SDR_a}$ y $\overline{SDR_{2s}}$ en azul, $\overline{SIR_o}$, $\overline{SIR_a}$ y $\overline{SIR_{2s}}$ en rojo, y $\overline{SAR_o}$, $\overline{SAR_a}$ y $\overline{SAR_{2s}}$ en negro. También aparecen los valores de los parámetros para la separación de 3 fuentes, $\overline{SDR_{3s}}$ en

azul claro, $\overline{SIR_{3s}}$ en granate y $\overline{SAR_{3s}}$ en gris.

Como se puede observar, la mejora entre el método de separación por onsets y por distancia armónica es más evidente en el valor de SIR, pero la diferencia numérica es bastante baja. Para valorarla adecuadamente conviene recordar que el conjunto de señales sobre el que se ha evaluado el método de distancias armónicas es más grande que el conjunto de señales analizadas por onsets. La diferencia en los resultados sonoros es más evidente. Las mejoras resultan mucho más palpables cuando entra en liza el método de separación de parciales superpuestos. Gracias a este, los resultados de distorsión y artefactos mejoran en 10dB aproximadamente, mientras que se mejora la relación señal a interferencia en más de 25dB.

De la pendiente positiva de estos datos se deduce la mejora paulatina en la calidad numérica. Del número de señales analizadas creciente y de la complejidad en aumento de las mezclas se infiere la robustez del método final. De la comparación acústica entre las señales originales y las separadas en cada uno de los tres casos, se desprenden posibles evoluciones futuras de esta aplicación.

Anexo IV

Tiempos de computación, f_{ins}, representaciones 2D y 3D

"No hay nada que limite más
la innovación que una visión dogmática
del mundo".
Stephen Jay Gould (1941-2002).
Paleontólogo y prominente
divulgador científico estadounidense.

En este Apéndice se completan los resultados experimentales relativos al Capítulo 5 de esta Tesis.

IV.a. Tiempos de computación: señales empleadas

El análisis de la velocidad de procesado del algoritmo CWAS se ha presentado en la Sección 5.2.3. De los resultados se deduce que el algoritmo presentado en esta disertación, pese a no haber sido optimizado, es entre 3 y 20 veces más rápido que algoritmos basados en la FFT (los cuales, a su vez, sí presentan una velocidad de proceso óptima). El estudio de los coeficientes de tiempo de procesado se ha llevado a cabo sobre un total de 25 señales de diferentes duraciones. Las especificaciones acerca de cada una de estas 25 señales se detallan en la Tabla IV.1.

En esta tabla aparecen los etiquetados empleados para cada señal en el texto (de s_1 a s_{25}), la señal correspondiente, y su duración en segundos y en número de muestras. Las

Señales empleadas en el análisis de tiempos de computación			
Nombre	**Señal**	**T (s)**	**L (m)**
s_1	*AltoFluteMfC5B5-11**	2.7856	61423
s_2	*AltosaxNoVibMfC4B4-3**	2.6554	58553
s_3	*BassFluteFfC3B3-1**	2.6202	57776
s_4	*BassoonFfC4B4-9**	2.2497	49606
s_5	*Clarinete*	1.1441	25228
s_6	*"Claros y frescos ríos"*	6.3668	140390
s_7	*Clean Guitar*	3.4738	76598
s_8	*Cuerdas (violines y violas) "nota"*	1.9992	44084
s_9	*Drum Kit (Batería)*	4.4322	97732
s_{10}	*Elvis "you"*	0.9412	20755
s_{11}	*Flauta C#5*	2.0493	45189
s_{12}	*Guitarra B4*	1.5819	34881
s_{13}	*Heaven and Hell Pt. 1-Excerpt*	36.4232	803133
s_{14}	*HornMfC4B4-5*	3.0000	66150
s_{15}	*Piano*	2.7928	61583
s_{16}	*Saxo*	1.7667	38956
s_{17}	*Tono puro 440Hz***	0.5000	22051
s_{18}	*TrombonBajoMfC3B3-7**	2.5364	55929
s_{19}	*TrombonBajoMfC3B3-10**	2.6000	57330
s_{20}	*TrompetaVibFfC5B5-9**	4.8000	105840
s_{21}	*ViolaArcoSulCFfC3B3-4**	2.7577	60809
s_{22}	*ViolaArcoSulCFfC3B3-5**	2.3056	50840
s_{23}	*Violin "corto"*	3.7667	83057
s_{24}	*Violin "largo"*	4.5538	100412
s_{25}	*Violín "nota"*	2.1094	46514

Tabla IV.1: *Información adicional sobre las 25 señales analizadas en la medida de tiempos de computación empleando la STFT y el algoritmo CWAS.*

señales marcadas con asterisco (*) provienen de la base de datos de instrumentos musicales de la Universidad de Iowa [63]. Todas las señales excepto el tono de 440Hz, s_{17}, marcado con dos asteriscos, (**) están muestreadas a 22050Hz.

IV.b. Acerca de la representación visual tiempo–frecuencia

En la Sección 5.4.1 se han presentado gráficas en dos y tres dimensiones de los espectrogramas obtenidos en el análisis de un grupo de diez señales (ocho sintéticas y dos reales) mediante algoritmos STFT, Wigner-Ville suavizada y empleando el algoritmo CWAS (bajo un número constante de $D = 24$ divisiones por octava). En el texto, se han mostrado tres de tales representaciones, concretamente las correspondientes a la señal mezcla de tres tonos puros y a las dos señales reales.

En las Figuras IV.1 y IV.2 se muestran los escalogramas bidimensionales y tridimensionales de cinco de las señales sintéticas, concretamente y por orden, el tono puro de 440Hz, la señal de FM, el chirp cuadrático, la señal mezcla de FM con tonos de 440Hz y 5kHz y la que se ha dado en llamar chirp "up-down" (creciente hiperbólico y descendente lineal). En cada figura, por columnas, aparecen los resultados obtenidos mediante la STFT, el algoritmo CWAS y la distribución de Wigner-Ville suavizada, respectivamente.

Como se puede comprobar, los espectrogramas obtenidos mediante el uso del algoritmo CWAS tienen tendencia a resultar más definidos en casi todo el espectro de frecuencias (la zona de sobreinformación por debajo de 100Hz que se hace evidente en las Figuras IV.2(h) y IV.2(n) se debe a que los filtros pasobanda de la parte baja del espectro son, con esta resolución, muy extensos en tiempo, lo que causa la deslocalización). Frente a la resolución constante de la STFT, la resolución variable del algoritmo CWAS supone una ventaja evidente. Por otro lado, en la última columna de estas figuras se puede apreciar como la WVD suavizada presenta un nivel de ruido bastante elevado, repartido por todo el espectro, además de componentes fantasma, producto de la interferencia entre los términos reales de la señal.

IV.c. Más datos sobre representaciones tiempo–frecuencia

En la Sección 5.3.3 se afirma que las técnicas de reasignación dentro de la Time Frequency Toolbox (TFTB) de Auger presentan una marcada tendencia a obtener un número de puntos mayor que la duración de la señal, cuando se trata de la PWVD, y un número de puntos marcadamente inferior en el caso de la RSP. Del mismo modo, el número de puntos obtenido al emplear las HRSR de Fulop (que depende básicamente del tamaño de la ventada de análisis) es, por otro lado, muy bajo. En la Tabla IV.2 se reflejan los resultados experimentales en los que se basan tales afirmaciones. En la tabla, los datos correspondientes a las señales marcadas con asterisco (*) han sido aproximados. La técnica de aproximación consiste en acotar la zona frecuencial de cada componente, obteniendo la cantidad de puntos de la transformada en esa zona.

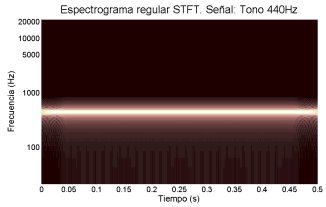
A partir de los datos experimentales de las HRSR, es posible sin embargo, aplicar una interpolación para obtener una frecuencia instantánea del tamaño adecuado, como se explica en la Sección 5.4.2.1.1. En cuanto a los resultados propios de las HRSR, tras la interpolación la tendencia es a la mejora en la estimación (aunque hay excepciones). En la Tabla IV.3 se detallan los resultados numéricos correspondientes. Los valores numéricos de precisión en la frecuencia instantánea interpolada son de una precisión muy elevada, tanto como puedan serlo los resultados del algoritmo CWAS. Estos datos son los que nos animaron a realizar el experimento de síntesis basado en la técnica de Fulop resumido en la Sección 5.4.2.1.1.

	Número de puntos obtenidos				
Señal	**Tamaño**	**RSP**	**RPWVD**	**HRSR**	**CWAS**
Tono 440	22051	18503	33503	172	22051
FM	22051	202	131495	172	22051
LC	22051	362	33153	172	22051
QC	22051	386	68558	172	22051
HC	22051	365	43632	172	22051
Chirp UD	44101	836	445703	346	44101
Tres tonos:*					
Tono 400Hz	44101	29790	38389	94	44101
Tono 600Hz	44101	29784	32337	94	44101
Tono 800Hz	44101	11737	37632	94	44101
FM+440+5k:*					
Tono 440Hz	44101	36908	47562	114	44101
FM	44101	31901	34635	94	44101
Tono 5kHz	44101	3275	10500	11	44101

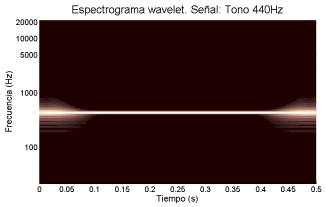
Tabla IV.2: *Número de puntos obtenidos en el estudio de las frecuencias instantáneas para cada una de las ocho señales sintéticas analizadas. En la columna 2 se refleja el tamaño real en muestras de cada señal. En las columnas 3 a 6, los resultados correspondientes a cada una de las herramientas de análisis empleadas, relacionadas con la reasignación, respectivamente, la RSP, RPWVD, HRSR y CWAS. Los datos correspondientes a las señales marcadas con asterisco (*) son aproximados.*

	Errores experimentales (%)		
Señal	**HRSR**	**HRSRi**	**CWAS**
Tono 440	0.1291	0.0672	0.0148
FM	0.0467	0.0402	0.0067
LC	0.0043	6.06E-4	0.0049
QC	0.211	0.2244	0.0172
HC	0.0097	0.0031	0.0099
Chirp UD	0.0173	0.0101	0.0306
Tres tonos:			
Tono 400Hz	0.6026	0.5767	4.93E-6
Tono 600Hz	0.6821	0.6594	3.29E-7
Tono 800Hz	0.9617	2.086	3.41E-5
FM+440+5k:			
Tono 440Hz	0.0367	0.0311	1.49E-5
FM	0.1862	0.279	2.68E-4
Tono 5kHz	0.0016	0.0016	6.77E-5

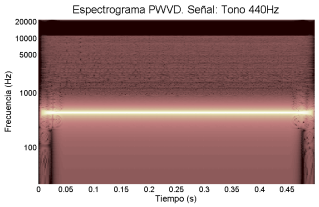
Tabla IV.3: *Comparativa de los errores en la detección de frecuencia instantánea cometidos empleando la HRSR antes y después de interpolar, y el algoritmo CWAS. Los datos aparecen en tanto por ciento (%).*



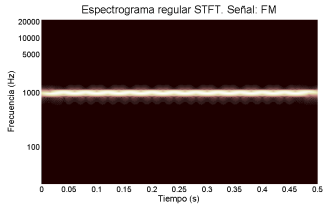
(a) *Espectrograma regular del tono puro de frecuencia 440Hz.*



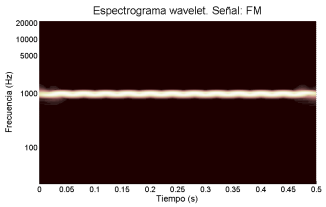
(b) *Espectrograma wavelet del tono puro de frecuencia 440Hz.*



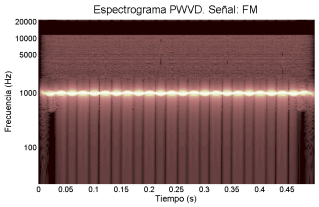
(c) *Espectrograma PWVD del tono puro de frecuencia 440Hz.*



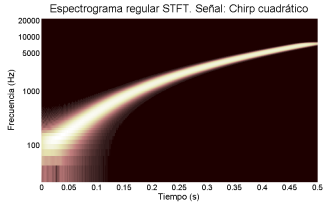
(d) *Espectrograma regular de la señal de FM.*



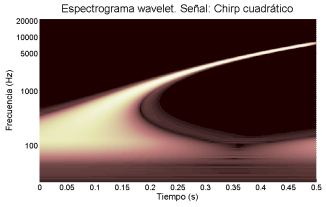
(e) *Espectrograma wavelet de la señal de FM.*



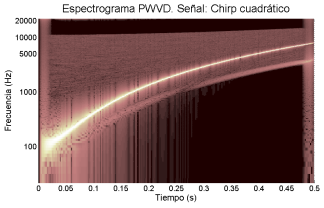
(f) *Espectrograma PWVD de la señal de FM.*



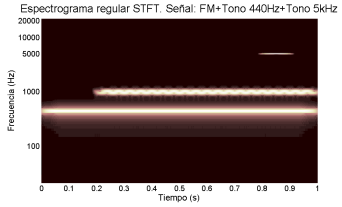
(g) *Espectrograma regular del chirp cuadrático.*



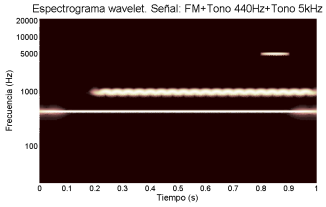
(h) *Espectrograma wavelet del chirp cuadrático.*



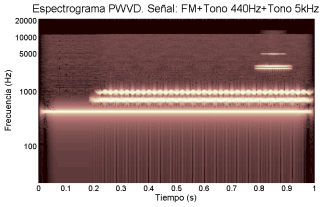
(i) *Espectrograma PWVD del chirp cuadrático.*



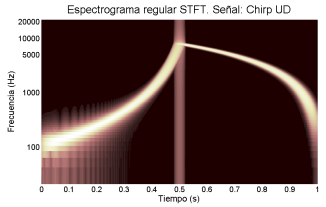
(j) *Espectrograma regular de la señal mezcla de FM y tonos de 440Hz y de 5kHz.*



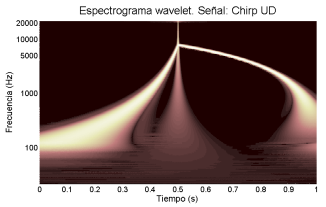
(k) *Espectrograma wavelet de la señal mezcla de FM y tonos de 440Hz y de 5kHz.*



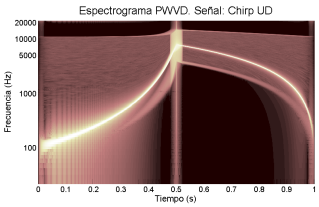
(l) *Espectrograma PWVD de la señal mezcla de FM y tonos de 440Hz y de 5kHz.*



(m) *Espectrograma regular de la señal de chirp UD.*

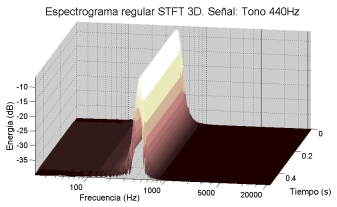


(n) *Espectrograma wavelet de la señal de chirp UD.*

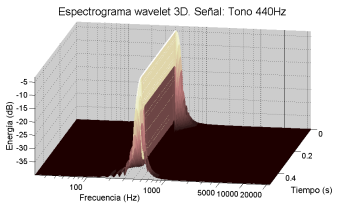


(ñ) *Espectrograma PWVD de la señal de chirp UD.*

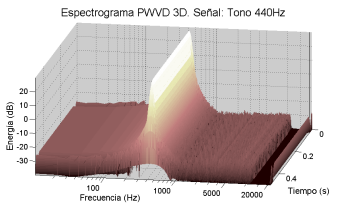
Figura IV.1: *Comparativa de los espectrogramas de 5 de las señales sintéticas analizadas. Por filas: el tono de 440Hz, la señal de FM, el chirp cuadrático, la mezcla de tono 440, señal de FM y tono de 5kHz y la señal chirp UD. (a), (d), (g), (j) y (m) Espectrogramas regulares obtenidos utilizando algoritmos FFT. (b), (e), (h), (k) y (n) Espectrogramas wavelet obtenidos mediante CWAS (D = 24 divisiones por octava). (c), (f), (i), (l) y (ñ) Espectrogramas obtenidos utilizando la transformada de Wigner-Ville suavizada.*



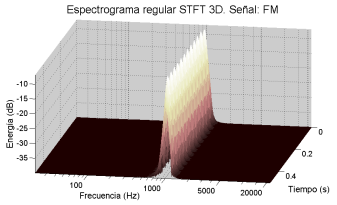
(a) *Espectrograma regular 3D del tono puro de frecuencia 440Hz.*



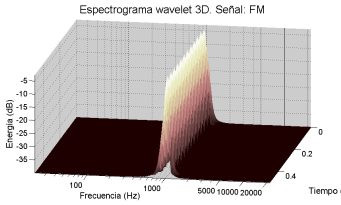
(b) *Espectrograma wavelet 3D del tono puro de frecuencia 440Hz.*



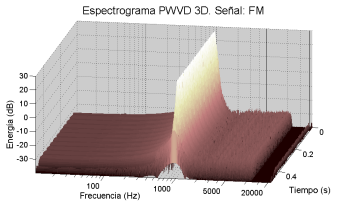
(c) *Espectrograma PWVD 3D del tono puro de frecuencia 440Hz.*



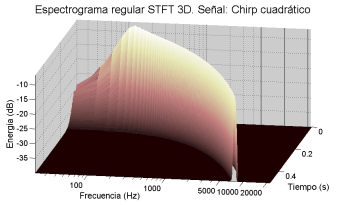
(d) *Espectrograma regular 3D de la señal de FM.*



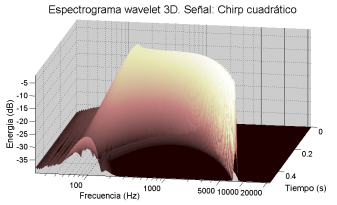
(e) *Espectrograma wavelet 3D de la señal de FM.*



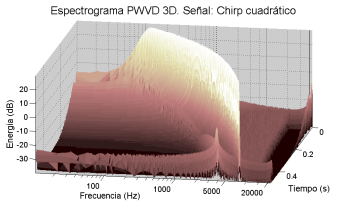
(f) *Espectrograma PWVD 3D de la señal de FM.*



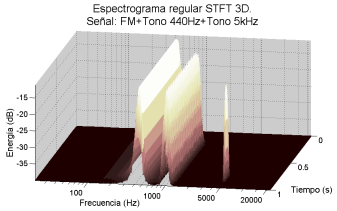
(g) *Espectrograma regular 3D del chirp cuadrático.*



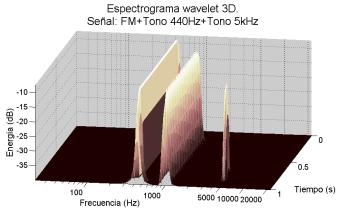
(h) *Espectrograma wavelet 3D del chirp cuadrático.*



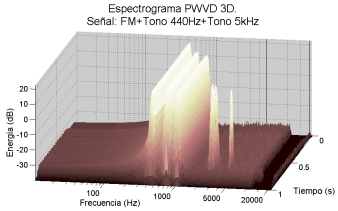
(i) *Espectrograma PWVD 3D del chirp cuadrático.*



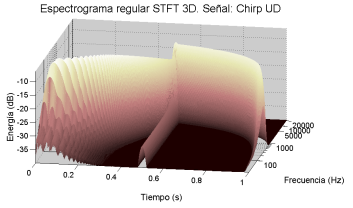
(j) *Espectrograma regular 3D de la señal mezcla de FM y tonos de 440Hz y de 5kHz.*



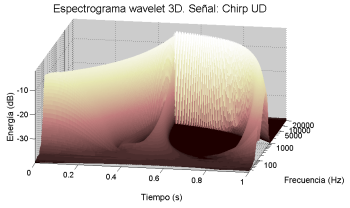
(k) *Espectrograma wavelet 3D de la señal mezcla de FM y tonos de 440Hz y de 5kHz.*



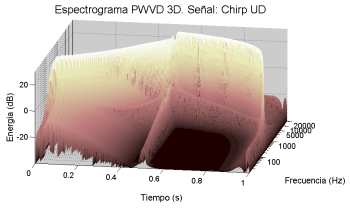
(l) *Espectrograma PWVD 3D de la señal mezcla de FM y tonos de 440Hz y de 5kHz.*



(m) *Espectrograma regular 3D de la señal de chirp UD.*



(n) *Espectrograma wavelet 3D de la señal de chirp UD.*



(ñ) *Espectrograma PWVD 3D de la señal de chirp UD.*

Figura IV.2: *Comparativa de los espectrogramas tridimensionales de 5 de las señales sintéticas analizadas. Por filas: el tono de 440Hz, la señal de FM, el chirp cuadrático, la mezcla de tono 440, señal de FM y tono de 5kHz y la señal chirp UD. (a), (d), (g), (j) y (m) Espectrogramas 3D regulares obtenidos utilizando algoritmos FFT. (b), (e), (h), (k) y (n) Espectrogramas wavelet 3D obtenidos mediante CWAS (D = 24 divisiones por octava). (c), (f), (i), (l) y (ñ) Espectrogramas 3D obtenidos utilizando la transformada de Wigner-Ville suavizada.*

Bibliografía

[1] S. A. Abdallah, *Towards Music Perception by Redundancy Reduction and Unsupervised Learning in Probablistic Models*, Ph.D. thesis, King's College London, 2002.

[2] P. S. Addison, *Wavelet transforms and the ECG: a review*, Physiological Measurement **26** (2005), R155–R199.

[3] A. N. Akansu, R. A. Haddad, and H. Caglar, *The Binomial QMF-Wavelet Transform for Multiresolution Signal Decomposition*, IEEE Transactions on Signal Processing **41, No. 1** (1993), 13–19.

[4] M. Akay and C. Mello, *Wavelets for Biomedical Signal Processing*, Proceedings of the 19th International Conference - IEEE/EMBS (1997), 2688–2691.

[5] S. Amari and J.F. Cardoso, *Blind Source Separation – Semiparametric Statistical Approach*, IEEE Transactions on Signal Processing **45, No. 11** (1997), 2692–2700.

[6] G. A. Arango, *Clasificación de fallas en motores eléctricos utilizando señales de vibración*, Tech. report, Universidad Tecnológica de Pereira, 2007.

[7] F. Auger, P. Flandrin, , P. Gonçalvès, and O. Lemoine, *Time–frequency toolbox*, [Online]. Available: http://tftb.nongnu.org, 1996.

[8] F. Auger and P. Flandrin, *Generalization of the reassignment method to all bilinear time-frequency and time-scale representations*, IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-94 **IV** (1994), 317–320.

[9] ———, *Improving the readability of time-frequency and time-scale representations by the reassignment method*, IEEE Transactions on Signal Processing **43, No. 5** (1995), 1068–1089.

[10] F. Auger, P. Flandrin, P. Gonçalvès, and O. Lemoine, *Time-Frequency Toolbox for use with MATLAB: Tutorial*, Tech. report, CNRS (France) and Rice University (USA), 1996.

[11] ———, *Time-Frequency Toolbox for use with MATLAB: Reference guide*, Tech. report, CNRS (France) and Rice University (USA), 1997.

[12] Varios autores, *Wikipedia. la enciclopedia libre*, [Online]. Available: http://es.wikipedia.org/wiki/, 2011.

[13] J. Backus, *The Acoustical Foundations of Music. Second Edition*, ch. 5. The Ear: Intensity and Loudness Levels, pp. 87–106, W.W. Norton and Company, New York, 1977.

[14] T. Bartosch and J. Wassermann, *Wavelet Coherence Analysis of Broadband Array Data Recorded at Stromboli Volcano, Italy*, Bulletin of the Seismological Society of America **94, No. 1** (2004), 44–52.

[15] E. Bedrosian, *A Product Theorem for Hilbert Transforms*, Proceedings of the IEEE **51** (1963), 868–869.

[16] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davis, and M. Sandler, *A Tutorial on Onset Detection in Music Signals*, IEEE Transactions on Speech and Audio Processing **13, No. 5** (September 2005), 1035–1047.

[17] J. R. Beltrán and F. Beltrán, *Additive Synthesis Based on the Continuous Wavelet Transform: a Sinusoidal Plus Transient Model*, Proc. of the 6th International Conference on Digital Audio Effects (DAFx-03) (2003).

[18] J. R. Beltrán and J. Ponce de León, *Blind Separation of Overlapping Partials in Harmonic Musical Notes Using Amplitude and Phase Reconstruction*, Submitted to EURASIP Journal on Advances in Signal Processing, Manuscript ID: 1711035958553557.

[19] ———, *Analysis and Synthesis of Sounds through Complex Bandpass Filterbanks*, Proc. of the 118th Convention of the Audio Engineering Society (AES'05). Preprint 6361 (2005).

[20] ———, *Extracción de Leyes de Variación Frecuenciales Mediante la Transformada Wavelet Continua Compleja*, Proceedings of the XX Simposium Nacional de la Unión Científica Internacional de Radio (URSI'05) (2005).

[21] ———, *Blind Source Separation of Monaural Musical Signals Using Complex Wavelets*, Proceedings of the 12th International Conference on Digital Audio Effects (DAFx-09) (2009).

[22] ———, *Estimation of the Instantaneous Amplitude and the Instantaneous Frequency of Audio Signals using Complex Wavelets*, Signal Processing **90/12** (2010), 3093–3109.

[23] J. R. Beltrán, J. Ponce de León, N. Degara, and A. Pena, *Localización de Onsets en Señales Musicales a través de Filtros Pasobanda Complejos*, Proceedings of the XXIII Simposium Nacional de la Unión Científica Internacional de Radio (URSI'08) (2008).

[24] J. R. Beltrán, J. Ponce de León, and E. Estopiñán, *Intermodulation Effects Analysis Using Complex Bandpass Filterbanks*, Proceedings of the 8th International Conference on Digital Audio Effects (DAFx-05) (2005), 149–154.

[25] J. R. Beltrán, *Detección y clasificación de contornos en imágenes mediante la Transformada Wavelet*, Ph.D. thesis, Dpt. of Electronic Engineering and Comunications, University of Zaragoza, 1994.

[26] J. R. Beltrán, F. Beltrán, and A. Estopañan, *Multiresolution edge clasification: Noise characterization*, Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics **5** (1998), 4476–4481.

[27] J. R. Beltrán, J. García-Lucía, and J. Navarro, *Edge detection and classification using Mallat's wavelet*, Proceeding of the IEEE International Conference on Image Processing **1** (1994), 293–296.

[28] J. R. Beltrán and J. Navarro, *Wavelet-based edge detection and classification*, Proceedings of EUSIPCO **September** (1994), 1381–1384.

[29] J. R. Beltrán and G. Palacios, *A wavelet transform based multiresolution edge detection and classification schema*, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 2008.

[30] B. Boashash, *Time-frequency signal analysis*, Advances in Spectrum Estimation and Array Processing (S. Haykin, ed.), vol. 1, Prentice Hall, Englewood Cliffs, NJ, 1990, pp. 418–517.

[31] ———, *Estimating and Interpreting the Instantaneous Frequency of a Signal. Part 1: Fundamentals*, Proceedings of the IEEE **80, No. 4** (1992), no. 4, 520–538.

[32] ———, *Estimating and Interpreting the Instantaneous Frequency of a Signal. Part2: Algorithms and Applications*, Proceedings of the IEEE **80, No. 4** (1992), no. 4, 540–568.

[33] B. Bogert, M. Healy, and J. W. Tukey, *The quefrency alanysis of time series for echoes: Cepstrum, pseudoautocovariance, cross-cepstrum, and saphe-cracking*, ch. 15, pp. 209–243, Proceedings of the Symposium on Time Series Analysis, 1963.

[34] A. S. Bregman, *Auditory Scene Analysis: The perceptual organization of sound*, MIT Press, 1990.

[35] F. Briz, *Cambio de afinación con preservación de timbre para señales de audio utilizando la transformada wavelet compleja*, Tech. report, Dept. Ingeniería Electrónica y Comunicaciones. Universidad de Zaragoza, 2006.

[36] F. Briz, M. W. Degner, P. García, and D. Bragado, *Broken Rotor Bar Detection in Line-Fed Induction Machines Using Complex Wavelet Analysis of Startup Transients*, IEEE Transactions on Industry Applications **44, No. 3** (2008), 760–768.

[37] G. J. Brown and M. Cooke, *Computational Auditory Scene Analysis*, Computer speech & language, Elsevier **8, No. 4** (1994), 297–336.

[38] J.J. Burred and T. Sikora, *On the Use of Auditory Representations for Sparsity-Based Sound Source Separation*, Fifth International Conference on Information, Communications and Signal Processing, ICICS05 (2005), 1466–1470.

[39] A. P. Calderón, *Intermediate spaces and interpolation, the complex method*, Studia Math. **24** (1964), 113–190. MR MR0167830 (29 #5097)

[40] J. F. Cardoso, *Blind signal separation: Statistical principles*, Proceedings of the IEEE **86** (October 1998), 2009–2025.

[41] R. A. Carmona, W. L. Hwang, and B. Torrésani, *Characterization of signals by the ridges of their wavelet transforms*, IEEE Transactions on Signal Processing **45** (1997), no. 10, 2586–2590.

[42] ———, *Multiridge detection and time-frequency reconstruction*, IEEE Transactions on Signal Processing **47** (1999), no. 2, 480–492.

[43] G. Cauwenberghs, *Monaural Separation of Independent Acoustical Components*, Proceedings of the 1999 IEEE International Symposium on Circuits and Systems-ISCAS '99 **5** (1999), 62–65.

[44] A. Chakraborty and D. Okaya, *Frequency-time decomposition of seismic data using wavelet-based methods*, Geophysics **60, No. 6** (1995), 1906–1916.

[45] E. Chassande-Mottin, F. Auger, and P. Flandrin, *Supervised time-frequency reassignment*, Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis (1996), 517–520.

[46] M. Cobos, *Application of Sound Source Separation Methods to Advanced Spatial Audio Systems*, Ph.D. thesis, Universidad Politécnica de Valencia, 2009.

[47] M. Cobos and J. J. López, *Stereo audio source separation based on time-frequency masking and multilevel thresholding*, Digital Signal Processing **18** (2008), 960–976.

[48] E. Copson, *Asymptotic expansions*, Cambridge Univ. Press, Cambridge, 1965.

[49] P. W. Crowley, *A Guide to Wavelets for Economists*, Journal of Economic Surveys **21, No. 2** (2007), 207–267.

[50] I. Daubechies, *Ten Lectures on wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 61, CBMS-NSF Series Appl. Math., SIAM, 1992.

[51] N. Delprat, B. Escudié, P. Guillemain, R. Kronland-Martinet, P. Tchamitchian, and B. Torrésani, *Asymptotic Wavelet and Gabor Analysis: Extraction of Instantaneous Frequencies*, IEEE Transactions on Information Theory **38** (1992), no. 2, 644–664.

[52] P. Depalle, G. Garcia, and X. Rodet, *Tracking of Partials for Additive Sound Synthesis using Hidden Markov Model*, IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'93 **1** (1993), 225–228.

[53] R. B. Dingle, *Asymptotic Expansions, their derivation and interpretation*, Academic Press, New York, 1973.

[54] Z. Duan, Y. Zhang, C. Zhang, and Z. Shi, *Unsupervised single-channel music source separation by average harmonic structure modeling*, IEEE Transactions on Audio, Speech and Language Processing **16, No. 4** (2008), 766–778.

[55] J. Ester, *Separación ciega de fuentes de audio en sonidos monaurales*, Tech. report, Centro Politécnico Superior, Universidad de Zaragoza, 2009.

[56] F. A. Everest, *Handbook for Sound Engineers. The New Audio Cyclopedia*, 2nd. ed., ch. 2. Psychoacoustics, pp. 25–42, SAMS, 1991.

[57] M. R. Every and J. E. Szymanski, *Separation of Synchronous Pitched Notes by Spectral Filtering of Harmonics*, IEEE Transactions on audio, speech and language processing **14, No. 5** (2006), 1845–1856.

[58] K. Fitz and S. A. Fulop, *A Unified Theory of Time-Frequency Reassignment*, Digital Signal Processing **arXiv:0903.3080** (2005).

[59] K. Fitz and L. Haken, *On the Use of Time–Frequency Reassignment in Additive Sound Modeling*, Journal of the Audio Engineering Society **50, No. 11** (2002), 879–893.

[60] K. Fitz, L. Haken, S. Lefvert, and M. O'Donnell, *Sound morphing using Loris and the reassigned bandwidth-enhanced additive sound model: Practice and applications*, International Computer Music Conference, (Gotenborg, Sweden) (2002).

[61] P. Flandrin, F. Auger, and E. Chassande-Mottin, *Applications in time-frequency signal processing*, ch. 5, Time-Frequency reassingnment: from principles to algorithms, pp. 179–204, CRC Press LLC, 2003.

[62] P. Frick, D. Galyagin, D. V. Hoyt, E. Nesme-Ribes, K. H. Schatten, D. Sokoloff, and V. Zakhzrov, *Wavelet analysis of solar activity recorded by sunspot groups*, Astronomy and Astrophysics **328, No. 2** (1997), 670–681.

[63] L. Fritts and Electronic Music Studios (University of Iowa), *Musical Instrument Samples Database*, [Online]. Available: http://theremin.music.uiowa.edu/MIS.html, (temporally out of service).

[64] S. Fulop, *High resolution spectrographic routines*, [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/21736-high-resolution-spectrographic-routines, 2008.

[65] S. A. Fulop and K. Fitz, *A Spectrogram for the Twenty-First Century*, Acoustics Today **July** (2006), 26–33.

[66] C. Févotte, R. Gribonval, and E. Vincent, *BSS EVAL Toolbox User Guide – Revision 2.0*, Tech. report, IRISA Technical Report 1706, Rennes, France, 2005.

[67] D. Gabor, *Theory of communication*, J. Inst. Elec. Eng. **93** (1946), no. III, 429–457.

[68] R. García Ramos, *Wavelets y su aplicación en el procesamiento de imágenes*, Ph.D. thesis, Universidad de las Américas Puebla, 2003.

[69] E. B. George and M. J. T. Smith, *Speech Analysis/Synthesis and Modification Using and Analysis-by-Synthesis/Overlap-Add Sinusoidal Model*, IEEE Transactions on Speech and Audio Processing **5, No. 5** (1997), 389–406.

[70] O. Gil, *Efectos de Audio Digitales Basados en la Transformada Wavelet Continua y Compleja*, Tech. report, Escuela Universitaria de Ingenieros Técnicos Industriales de Zaragoza. Universidad de Zaragoza, Previsto Septiembre 2011.

[71] R. Gribonval, E. Vincent, C. Févotte, and L. Benaroya, *Proposals for performance measurement in source separation*, Proceedings of the International Conference on Independent Component Analysis and Blind Source Separation (ICA) (2003), 763–768.

[72] A. Grossmann, R. Kronland-Martinet, and J. Morlet, *Reading and understanding the continuous wavelet transform*, Wavelets: Time-Frequency methods and phase space (J.M. Combes, A. Grossmann, and Ph.Tchamitchian, eds.), Springer, Berlin, 1989, pp. 2–20.

[73] A. Grossmann and J. Morlet, *Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape*, Siam Journal On Mathematical Analysis **15** (1984), 723–736.

[74] A. Grossmann, J. Morlet, and T. Paul, *Transforms Associated to Square Integrable Group-Representations .1. General Results*, Journal of Mathematical Physics **26** (1985), 2473–2479.

[75] _____, *Transforms Associated to Square Integrable Group-Representations .2. Examples*, Annales de L'Institut Henri Poincare-Physique Theorique **45** (1986), 293–309.

[76] P. Guillemain and R. Kronland-Martinet, *Characterization of Acoustic Signals through Continuous Linear Time-Frequency Representations*, Proceedings of the IEEE **84** (1996), no. 4, 561–585.

[77] S. L. Hahn, *Short Communication on the Uniqueness of the Definition of the Amplitude and Phase of the Analytic Signal*, Signal Processing **83** (2003), 1815–1820.

[78] S. W. Hainsworth and P. J. Wolfe, *Time–Frequency Reassignment for Music Analysis*, Proceedings of International Computer Music Conference ICMC'01 (2001), 12–17.

[79] J. Han and B. Pardo, *Reconstructing completely overlapped notes from musical mixtures*, Proceedings of the IEEE International Conference on Acoustics, Speech ans Signal Processing (ICASSP'11). (2011), 249–252.

[80] A. Hanss, T. Leifsen, B. N. Andersen, and T. Toutain, *Wavelet decomposition of global solar oscillations*, Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis (1994), 152–153.

[81] N. Hazarika, J. Zhu Chen, A. Chung Tsoi, and A. Sergejew, *Classification of EEG signals using the wavelet transform*, Signal Processing **59** (1997), 61–72.

[82] F. Hlawatsch and P. Flandrin, *The interference structure of the Wigner distribution and related time-frequency signal representations*, The Wigner distribution. Theory and applications in Signal Processing (W. F. G. Mecklenbrauker, ed.), Elsevier, Amsterdam, Amsterdam, 1993, pp. 59–133. MR 1 643 942

[83] E. Hostalkova and A. Prochazka, *Complex wavelet transform in biomedical image denoising*, Proceedings of 15th Annual Conference Technical Computing, Prague **HP** (2007), 1–8.

[84] G. Hu and D. Wang, *Monaural Speech Segregation Based on Pitch Tracking and Amplitude Modulation*, IEEE Transactions On Neural Networks **15, No. 5** (2004), 1135–1150.

[85] M. G. Jafari, S. A. Abdallah, M. D. Plumbey, and M. E. Davies, *Sparse Coding for Convolutive Blind Audio Source Separation*, Lecture Notes in Computer Science-Independent Component Analysis and Blind Signal Separation **3889** (2006), 132–139.

[86] D. Jiménez, *Análisis de señales ECG a través de wavelets complejas*, Tech. report, Escuela Universitaria de Ingenieros Técnicos Industriales de Zaragoza. Universidad de Zaragoza, Previsto 2011.

[87] I. Kauppinen, *Methods for detecting impulsive noise in speech and audio signals*, Proceedings of the 14th International Conference of Digital Signal Process (DSP-2002) **24** (2002), 967–970.

[88] N. Kingsbury, *Complex wavelets for shift invariant analysis and filtering of signals*, Journal of Applied and Computational Harmonic Analisys **10** (2001), 234–253.

[89] A. Klapuri, *Sound Onset detection by appliyng Phsycoacoustic Knowledge*, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'99 (1999), 115–118.

[90] ———, *Multipitch Estimation and Sound Separation by the Spectral Smoothness Principle*, Proceedings of the IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP'01) (2001), 3381–3384.

[91] ———, *Multiple Fundamental Frequency Estimation Based on Harmonicity and Spectral Smoothness*, IEEE Transactions on Speech and Audio Processing **11, No. 6** (2003), 804–816.

[92] K. Kodera, C. De Villedary, and R. Gendrin, *A new method for the numerical analysis of non-stationary signals*, Physics of The Earth and Planetary Interiors **12** (1976), 142–150.

[93] K. Kodera, R. Gendrin, and C. De Villedary, *Analysis of time-varying signals with small BT values*, IEEE Transactions on Acoustics, Speech and Signal Processing **ASSP-26, No. I** (1978), 64–76.

[94] E. D. Kolaczyk, *Methods for analyzing certain signals and images in astronomy using Haar wavelets*, Conference Record of the Thirty-First Asilomar Conference on Signals, Systems & Computers **1** (1997), 80–84.

[95] P. Kotsas, C. Pappas, M. Strintzis, and N. Maglaveras, *Nonstationary ECG Analysis Using Wigner-Ville Transform and Wavelets*, Computers in Cardiology (1993), 499–502.

[96] J. Kovacevic and M. Vetterli, *Perfect Reconstruction Filter Banks with Rational Sampling Rate Changes in One and Two Dimensions*, SPIE, Visual Communications and Image Processing **1199** (1989), 1258–1268.

[97] ———, *Perfect Reconstruction Filter Banks with Rational Sampling Rate Changes*, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '91) **May** (1991), 1785–1788.

[98] ———, *Design of Multidimensional Non-separable Regular Filter Banks and Wavelets*, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '92) **IV** (1992), 389–392.

[99] ———, *Nonseparable Multidimensional Perfect Reconstruction Filter Banks and Wavelet Bases for Rn*, IEEE Transactions on Information Theory **38, No. 2** (1992), 533–555.

[100] R. Kronland-Martinet, *The Wavelet Transform for Analysis, Synthesis and Processing of Speech and Music Sounds*, Computer Music Journal **12, No. 4** (1988), no. 4, 11–20.

[101] R. Kronland-Martinet, J. Morlet, and A. Grossmann, *Analysis of sound patterns through wavelet transforms*, International Journal of Pattern Recognition and Artificial Intelligence **1** (1987), no. 2, 272–302.

[102] M. Kulesh, M. Holschneider, and M. S. Diallo, *Geophysical wavelet library: Applications of the continuous wavelet transform to the polarization and dispersion analysis of signals*, Computers & Geosciences **34** (2008), 1732–1752.

[103] P. Kumar and E. Foufoula-Georgiou, *Wavelet analysis for geophysical applications*, Reviews of Geophysics **35, No. 4** (1997), 385–412.

[104] M. Lagrange, S. Marchand, M. Raspaud, and J.-B. Rault, *Enhanced Partial Tracking using Linear Prediction*, Proceedings of the 6th International Conference on Digital Audio Effects (DAFx-03) (2003), 141–146.

[105] M. Lagrange, S. Marchand, and J.-B. Rault, *Using linear prediction to enhance the tracking of partials*, IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'04 **4** (2004), 241–244.

[106] W. D. Langer, R. W. Wilson, and C. H. Anderson, *Hierarchical structure analysis of interstellar clouds using nonorthogonal wavelets*, Astrophysical Journal **408** (1993), L45–L48.

[107] J. Laroche and M. Dolson, *Improved phase vocoder time-scale modification of audio*, IEEE Transactions on Speech and Audio Processing **7** (1999), no. 3, 323–332.

[108] D. D. Lee and H. S. Seung, *Learning the Parts of Objects by Nonnegative Matrix Factorization*, Nature **401** (1999), 788–791.

[109] C. Li, C. Zheng, and C. Tai, *Detection of ECG Characteristic Points Using Wavelet Transforms*, IEEE Transactions on BiomedicaI Engineering **42, No. 1** (1995), 21–28.

[110] Y. Li, J. Woodruff, and D. Wang, *Monaural Musical Sound Separation Based on Pitch and Common Amplitude Modulation*, Transactions on Audio, Speech and Language Processing **17, No. 7** (2009), 1361–1371.

[111] S. Mallat, *A Theory for Multiresolution Signal Decomposition: The Wavelet Representation*, IEEE Tran. on Patt. Anal. and Machine Intell. **11** (1989), no. 7, 674–693.

[112] ———, *A Wavelet Tour of Signal Processing*, 2nd. ed., ch. 4, Time Meets Frequency, pp. 67–124, Academic Press, 1998.

[113] S. Mallat and S. Zhong, *Characterization of Signals from Multiscale Edges*, IEEE Transactions on Pattern Analysis and Machine Intelligence **14, N. 7** (1992), 710–732.

[114] S. Mann and S. Haykin, *The Chirplet Transform: Physical Considerations*, IEEE Transactions on Signal Processing **43, No. 11** (1995), 2745–2761.

[115] The MathWorks community, *Mathworks: MatlabCentral*, [Online]. Available: http://www.mathworks.com/matlabcentral/, 1994–2011.

[116] R. J. McAulay and T. F. Quatieri, *Speech Analysis/Synthesis Based on a Sinusoidal Representation*, IEEE Transaction on Acoustics, Speech and Signal Processing **34, no. 4** (1986), 744–754.

[117] T. Melia, *Underdetermined Blind Source Separation in Echoic Environments Using Linear Arrays and Sparse Representations*, Ph.D. thesis, School of Electrical, Electronic and Mechanical Engineering University College Dublin, National University of Ireland, 2007.

[118] M. Miller, N. Kingsbury, and R. W. Hobbs, *Seismic Imaging Using Complex Wavelets*, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '05) **2** (2005), 557–560.

[119] E. Moulines, F. Emerard, D. Larreur, J.L. Le Saint Milon, L. Le Faucheur, F. Marty, F. Charpentier, and C. Sorin, *A real-time French text-to-speech system generating high-quality synthetic speech*, International Conference on Acoustics, Speech, and Signal Processing, 1990. ICASSP-90. **1.** (1990), 309–312.

[120] A. Muñoz, R. Ertlé, and M. Unser, *Continuous Wavelet Transform with Arbitrary Scales and O(N) Complexity*, Signal Processing **82** (2002), 749–757.

[121] D. J. Nelson, *Cross-spectral methods for processing speech*, Journal of the Acoustical Society of America **110, No. 5** (2001), 2575–2592.

[122] ———, *Instantaneous Higher Order Phase Derivatives*, Digital Signal Processing **12** (2002), 416–428.

[123] D. E. Newland, *Practical signal analysis: Do wavelet make any difference?*, ASME Design Engineering Technical Conferences **DETC97/VIB-4135** (1997), 1–12.

[124] W. Nho and P. J. Loughlin, *When is instantaneous frequency the average frequency at each time?*, IEEE Signal Processing Letters **6, No. 4** (April 1999), no. 4, 78–80.

[125] P. M. Oliveira and V. Barroso, *Instantaneous frequency of multicomponent signals*, IEEE Signal Processing Letters **6, No. 4** (April 1999), no. 4, 81–83.

[126] A. V. Oppenheim and R. W. Schafer, *Discrete-time signal processing*, Prentice-Hall, 1989.

[127] L. I. Ortiz-Berenguer, *Identificación Automática de Acordes Musicales*, Ph.D. thesis, Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad Politécnica de Madrid, 2002.

[128] L. I. Ortiz-Berenguer, F. J. Casajús-Quirós, M. Torres-Guijarro, and J. A. Beracoechea, *Piano Transcription Using Pattern Recognition: Aspects on Parameter Extraction*, Proceedings of the 7th Conference on Digital Audio Effects (DAFx'04) (2004), 212–216.

[129] T. W. Parsons, *Separation of Speech from Interfering Speech by means of Harmonic Selection*, The Journal of the Acoustical Society of America **60, No. 4** (1976), 911–918.

[130] Z. K. Peng and F. L. Chu, *Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography*, Mechanical Systems and Signal Processing **18** (2004), 199–221.

[131] B. Picinbono and W. Martin, *Signal Representation by Instantaneous Amplitude and Phase*, Annales des Telecommunications-Annals Of Telecommunications **38** (1983), 179–190.

[132] T. F. Quatieri and R. J. McAulay, *Audio Signal Processing based on sinusoidal analysis/synthesis*, Applications of Digital Signal Processing to Audio and Acoustics (M. Kahrs and K. Brandenburg, eds.), Kluwer Academic Publishers, 1998.

[133] J. Ramsey and C. Lampart, *Decomposition of economic relationships by timescale using wavelets*, Macroeconomic Dynamics **2** (1998), 49–71.

[134] J. B. Ramsey, *Wavelets in Economics and Finance: Past and Future*, Studies in Nonlinear Dynamics & Econometrics **6, No. 3, Article 1** (2002), 1–27.

[135] S. Rickard, *Blind Speech Separation*, ch. 8. The DUET Blind Source Separation Algorithm, pp. 217–241, Springer Netherlands, 2007.

[136] A. Rihaczek, *Hilbert Transforms and the Complex Representations of Signals*, Proc. of the IEEE **54** (1966), 434–435.

[137] ———, *Signal Energy Distribution In Time And Frequency*, IEEE Transactions on Information Theory **IT-14, No. 3** (1968), 369–374.

[138] O. Rioul and M. Vetterli, *Wavelets and Signal Processing*, IEEE Signal Processing Magazine **8, n.4** (1991), 14–38.

[139] X. Rodet, *Time-domain formant-wave-function synthesis*, Spoken Language Generation and Understanding (J. G. Simon, ed.), Dordrecht: D. Reidel, 1980.

[140] X. Rodet and Ph. Depalle, *A new additive synthesis method using inverse Fourier transform and spectral envelopes*, Proceedings of the International Computer Music Conference, ICMC'92, October 1992.

[141] J. S. Sahambi, S. N. Tandon, and R. K. P. Bhatt, *Using Wavelet Transforms for ECG Characterization. An on-line digital signal processing system*, IEEE Engineering in Medicine and Biology **Jan/Feb** (1997), 77–83.

[142] G. Saracco, P. Labazuy, and F. Moreau, *Localization of self-potential sources in volcano-electric effect with complex continuous wavelet transform and electrical tomography methods for an active volcano*, Geophysical Research Letters **31** (2004).

[143] M. N. Schmidt and R. K. Olsson, *Single-channel Speech Separation Using Sparse Non-negative Matrix Factorization*, Internationnal Conference on Spoken Languaje Processing, ICSLP'06 (2006), 2614–2617.

[144] N. M. Schmidt and Mørup M., *Nonnegative Matrix Factor 2-D Deconvolution for Blind Single Channel Source Separation*, Proceedings of the 6th International Conference on

Independent Component Analysis and Blind Signal Separation, ICA'06, Lecture Notes in Computer Science **3889** (2006), 700–707.

[145] A. Schuck Jr. and J. 0. Wisbeck, *QRS Detector Pre-processing Using the Complex Wavelet Transform*, Proceedings of the 25th Annual Intemational Conference of the IEEE Engineering in Medicine and Biology Society (2003), 2590–2593.

[146] I. W. Selesnick, *Hilbert Transform Pairs of Wavelet Bases*, Signal Processing Letters **8, No. 6** (2001), 170–173.

[147] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, *The Dual-Tree Complex Wavelet Transform*, IEEE Signal Processing Magazine **November** (2005), 123–151.

[148] X. Serra, *A System for Sound Analysis/Transformation/Synthesis based on a Deterministic plus Stochastic Decomposition*, Ph.D. thesis, Stanford University, 1989.

[149] ———, *Residual Minimization in a Musical Signal Model based on a Deterministic plus Stochastic Decomposition*, Journal of the Acoustical Society of America **95, (5-2)** (1994), 2958–2959.

[150] ———, *Musical Signal Processing*, ch. Musical Sound Modeling with Sinusoids plus Noise, pp. 91–122, Swets & Zeitlinger, 1997.

[151] ———, *Spectral Modeling Synthesis: Past and Present*, keynote at DAFx 2003, London, September 10th 2003, 2003.

[152] X. Serra, J. Bonada, P. Herrera, and R. Loureiro, *Integrating complementary spectral models in the design of a musical synthesizer*, Proceedings of the International Computer Music Conference **September** (1997), 152–159.

[153] X. Serra and Music Technology Group (Universitat Pompeu Fabra), *The Clam Project*, [Online]. Available: http://clam-project.org/download.html, 2009.

[154] X. Serra and J. O. Smith, *Spectral modelling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition*, Computer Music Journal **14** (1990), no. 4, 12–24.

[155] P. D. Shukla, *Complex Wavelet Transforms and Their Applications*, Ph.D. thesis, The University of Strathclyde in Glasgow, 2003.

[156] D. Slepian, *On Bandwidth*, Proceedings of the IEEE **64** (1976), 292–300.

[157] D. Slepian, H. O. Pollak, and H. J. Landau, *Prolate spheroidal wave functions, Fourier analysis and uncertainty-I, II*, Bell Syst. Tech. J. **40** (1961), 43–83.

[158] R. Spaendonck, T. Blu, R. Baraniuk, and M. Vetterli, *Orthogonal Hilbert Transform Filter Banks and Wavelets*, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '03). **VI** (2003), 505–508.

[159] M. R. Spiegel, *Manual de fórmulas y tablas matemáticas (colección Schaum)*, McGraw Hill/Interamericana de México, 1968-1991.

[160] J.-L. Starck and J. Bobin, *Astronomical Data Analysis and Sparsity: From Wavelets to Compressed Sensing*, Proceedings of the IEEE **98** (2009), 1021–1030.

[161] J.-L. Starck and F. Murtagh, *Handbook of astronomical data analysis*, Springer, 2006.

[162] C. Torrence and G. P. Compo, *A Practical Guide to Wavelet Analysis*, Bulletin of the American Meteorological Society **79, No. 1** (1998), 61–78.

[163] M. Vetterli and J. Kovacevic, *Perfect Reconstruction Filter Bnaks for HDTV Representation and Coding*, Elsevier Signal Processing: Image Communication **2** (1990), 349–363.

[164] J. Ville, *Theorie et applications de la notion de signal analytique*, Cables et Transm. **2A** (1948), no. 10, 61–74.

[165] E. Vincent, R. Gribonval, and C. Févotte, *Performance Measurement in Blind Audio Source Separation*, IEEE Transactions on Audio, Speech and Language Processing **14, No. 4** (2006), 1462–1469.

[166] T. Virtanen, *Sound Source Separation in Monaural Music Signals*, Ph.D. thesis, Tampere University of Technology, 2006.

[167] ———, *Monaural Sound Source Separation by Non-Negative Matrix Factorization with Temporal Continuity and Sparseness Criteria*, IEEE Transactions on Audio, Speech and Language Processing **15, No. 3** (2007), 1066–1074.

[168] T. Virtanen and A. Klapuri, *Separation of Harmonic Sound Sources Using Sinusoidal Modeling*, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP '00 **2** (2000), 765–768.

[169] ———, *Separation of Harmonic Sounds Using Multipitch Analysis and Iterative Parameter Estimation*, IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (2001), 83–86.

[170] D. Wang and G. J. Brown, *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*, Wiley-IEEE Press, 2006.

[171] E. P. Wigner, *On the Quantum Correction For Thermodynamic Equilibrium*, Physical Review **40** (1932), no. 5, 749–759.

[172] R. Willet, I. Jermyn, R. Nowak, and J. Zerubia, *Wavelet-Based Superresolution in Astronomy*, Proceedings of Astronomical Data Analysis Software and Systems (Astronomical Society of the Pacific) **314** (2004).

[173] J. Woodruff, Y. Li, and D. Wang, *Resolving Overlapping Harmonics for Monaural Musical Sound Separation Using Pitch and Common Amplitude Modulation*, Proceedings of the International Conference on Music Information Retrieval (2008), 538–543.

[174] J. Xiao and P. Flandrin, *Multitaper time-frequency reassignment for nonstationary spectrum estimation and chirp enhancement*, IEEE Transactions on Signal Processing **55, No. 6** (2007), 2851–2860.

[175] Ö. Yilmaz and S. Rickard, *Blind separation of speech mixtures via time-frequency masking*, IEEE Transactions On Signal Processing **52, No. 7** (2004), 1830–1847.

[176] X. Zhou-min, W. En-fu, Z. Guo-hong, Z. Guo-cun, and C. Xu-geng, *Seismic signal analysis based on the dual-tree complex wavelet packet transform*, Acta Seismologica Sinica **17** (2004), 117–122.

[177] M. Zivanovic, *Detection of non-stationary sinusoids by using joint frequency reassignment and null-to-null bandwidth*, Digital Signal Processing **doi:10.1016/j.dsp.2010.03.011** (2010).

[178] U. Zölzer (ed.), *DAFX. Digital Audio Effects*, John Wiley & sons, Ltd., 2002.

Publicaciones

Artículos publicados por el autor en revistas indexadas en el JCR, congresos internacionales de reconocido prestigio y congresos nacionales, ya sea como autor principal (AP) o segundo autor (CA) siendo el primer autor el director de la Tesis *"Análisis y Síntesis de Señales de Audio a través de la Transformada Wavelet Continua y Compleja: El Algoritmo CWAS"*.

Autor: *Jesús Ponce de León Vázquez.*

Director: *José Ramón Beltrán Blázquez.*

Publicaciones en revistas indexadas en el JCR:

1. **Estimation of the Instantaneous Amplitude and the Instantaneous Frequency of Audio Signals using Complex Wavelets**
   José Ramón Beltrán y Jesús Ponce de León.
   Signal Processing, **2010**, *90/12*, 3093–3109. **CA**.

2. **Blind Separation of Overlapping Partials in Harmonic Musical Notes Using Amplitude and Phase Reconstruction**
   Jesús Ponce de León y José Ramón Beltrán.
   EURASIP Journal on Advances in Signal Processing, Manuscript ID: 1711035958553557, **pendiente de aceptación**. **AP**.

Publicaciones en Congresos Nacionales:

3. **Extracción de Leyes de Variación Frecuenciales Mediante la Transformada Wavelet Continua Compleja**
   José Ramón Beltrán y Jesús Ponce de León.
   Proceedings of the XX Simposium Nacional de la Unión Científica Internacional de Radio (URSI'05), **2005**. **CA**.

4. **Localización de Onsets en Señales Musicales a través de Filtros Pasobanda Complejos**
   José Ramón Beltrán, Jesús Ponce de León, Norberto Degara y Antonio Pena.
   Proceedings of the XXIII Simposium Nacional de la Unión Científica Internacional de Radio (URSI'08), **2008**. **CA**.

Publicaciones en Congresos Internacionales:

5. **Analysis and Synthesis of Sounds through Complex Bandpass Filterbanks**
   José Ramón Beltrán y Jesús Ponce de León.
   Proceedings of the 118^{th} *Convention of the Audio Engineering Society (AES'05)*, **2005**, preprint 6361. **CA**.

6. **Intermodulation Effects Analysis Using Complex Bandpass Filterbanks**
   José Ramón Beltrán, Jesús Ponce de León y Eduardo Estopiñán.
   Proceedings of the 8^{th} *International Conference on Digital Audio Effects (DAFx-05)* , **2005**, 149–154. **CA**.

7. **Blind Source Separation of Monaural Musical Signals Using Complex Wavelets**
   José Ramón Beltrán y Jesús Ponce de León.
   Proceedings of the 12^{th} *International Conference on Digital Audio Effects (DAFx-09)*, **2009**. **CA**.